National Library of Canada

Acquisitions and Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada

Direction des acquisitions et des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

THE UNIVERSITY OF ALBERTA

X-ray Crystallographic Studies on DNA Binding Antibodies

BY

Clifford Dean Mol

©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Biochemistry
Edmonton, Alberta
SPRING 1993

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-82146-9

Canada

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Clifford Dean Mol

TITLE OF THESIS: X-ray Crystallographic Studies on DNA Binding Antibodies

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1992

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Clifford Mol

3503 148 Avenue
Edmonton, alberta
Canada, T5Y 2M2

December 31, 1992

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **X-ray Crystallographic Studies on DNA Binding Antibodies** submitted by **Clifford Dean Mol** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**

Wayne F. Anderson
Dr. W.F. Anderson

Michael James
Dr. M.N.G. James

A.R. Morgan
Dr. A.R. Morgan

Dr. W.A. Bridger

Steven Sheriff
Dr. S. Sheriff

Linda Pilarski
Dr. L. Pilarski

November 20, 1992

Dedicated to my mother, Jerina Mol, for her love and support

ABSTRACT

Nucleic acid binding antibodies are an interesting model system for studying protein-nucleic acid interactions. The ability to obtain large quantities of a single homogenous antibody preparation, and the relative ease with which the antigen binding (Fab) fragments can be prepared and crystallized, makes this system amenable to study by x-ray crystallographic methods. Furthermore, the solution of Fab fragment structures can be achieved in a straightforward manner using the techniques of molecular replacement, which takes advantage of the fact that a large portion of the structure of an antibody Fab fragment is conserved. For these reasons we have chosen to determine the structures of nucleic acid binding antibodies by x-ray crystallography.

It has been found that the isoelectric purity of the Fab fragment preparations can affect the growth of crystals suitable for diffraction analyses. Purification of the Fab fragments to isoelectric homogeneity greatly facilitated the growth of large single crystals. The purification of the Fab fragments from two murine monoclonal DNA binding antibodies, and the preparation of suitable crystals, is described. These antibodies are Jel 72, which is specific for the right handed DNA duplex poly(dG)·poly(dC), and Jel 318, which binds triplex DNA with the sequence poly(dTm^{5}C)·poly(dGA)·poly(dTm^{5}C$^+$).

The structure determination and crystallographic refinements of two crystal forms of Fab Jel 72 and one crystal form of Fab Jel 318 are described. Jel 72 (form I) crystallizes in space group $P2_{1}2_{1}2_{1}$, a=94.63, b=102.60, c=92.42 Å, and has been refined to an R value of 0.18 for 19,136 reflections at 2.7 Å resolution. Jel 72 (form II) also crystallizes in space group $P2_{1}2_{1}2_{1}$, a=90.56, b=140.71, c=37.44 Å, and has been refined to an R value of 0.23 for 6,234 reflections at 2.9 Å resolution. Jel 318, space group $P2_{1}2_{1}2_{1}$, a=81.89, b=139.19, c=40.96 Å, has been refined to an R value of 0.20 for 9,342 reflections at a resolution of 2.8 Å.

Models for the antibody combining regions were constructed and are compared to the observed crystal structures. The results indicate that reasonable models for antibody combining regions can be obtained using a template-based modelling procedure. The modelling protocol was then applied to three other nucleic acid binding antibodies whose three dimensional structures have not yet been determined.

An examination of the shapes and charge distributions of the antibody combining sites is used to propose working models for the interaction of these antibodies with their respective nucleic acid antigens. These models may prove useful in the design of, and interpretation of results from, biochemical and immunological experiments to further elucidate the mechanism of DNA binding by these antibodies.

Acknowledgement

I acknowledge the advice and guidance of my supervisor, Wayne F. Anderson, and his wife Caroline, throughout the past several years, which have seen me go from Edmonton to Nashville, back to Edmonton, back to Nashville, and finally back to Edmonton, again . I also wish to acknowledge the contributions of many friends and colleagues who have worked with me in Wayne's laboratories or in neighbouring laboratories, both at the University of Alberta and at Vanderbilt University in Nashville, Tennessee where we relocated in 1988-89. These people include; Mr. Amechand Boodhoo, Mr. David Bacon, Dr. Mirek Cygler, Mr. Eric Harley, Dr. Alastair Muir, and Ms. Grace Ozimek, in Edmonton, the first time.

The people I have met in Tennessee are among the friendliest and kindest people I have ever had the pleasure to know. They include: Dr. Steve Geib, Dr. Marcia Newcomer, Dr. Steve Pappas, Mr. D. Brian Prince, Celeste and Danny Riley, Dr. Udai Singh, Ms. Zena Weeks, and Dr. Alan Wilcox, in Nashville, the first time.

The people whose advice and guidance helped me during my year long 'sabbatical' in Edmonton include Mr. Steve Mosiman and Mr. Russell Bishop, as well as Dr. Randy Read, Dr. Trevor Hart, Dr. W.A. Bridger, Dr. M.N.G. James, and Dr. A.R. Morgan. Upon my return to Nashville the second time, I am forever indebted to the helpful advice and guidance of Celeste and

God smile on you both.

Throughout my graduate work I have been fortunate to have the love and support of my family and friends. I would like to particularly acknowledge the love and emotional, as well as financial, support I have received from my mother, Jerina Mol. As well, my younger brother, Rodney, and my older brother, Alex, and my good friend, John Shopsky, have been steadfast in their support of my education, even though at times they must have thought I was crazy.

This research has been supported by the Medical Research Council of Canada, through a grant to the MRC Group in Protein Structure and Function, by the Alberta Heritage foundation, through a Studentship awarded to me, by a grant from the National Institutes of Health, and by the Student Finance Board in Edmonton, through a student loan awarded to me.

Table of Contents

Chapter page

I. **Introduction**

I.1 Anti-DNA Antibodies........................1
I.2 Structure of Immunoglobulin G................9
I.3 Antibody-Antigen Complexes..................24
McPC603 and phosphocholine..................24
4-4-20 and fluorescein......................25
BV04-01 and $d(pT)_3$.......................27
The Anti-Lysozyme complexes.................33
I.4 DNA Binding Proteins........................36
The Helix-turn-Helix Motif..................36
EcoR1 Endonuclease..........................39
Zinc Finger Proteins........................40
Proteins that use ß-strands to bind DNA.....41
I.5 Conclusions.................................43

II. **Purification and Crystallization**

II.1 Introduction...............................47
II.2 Purification of Jel 72.....................49
II.3 Crystallization of Jel 72 (form I).........54
II.4 Crystallization of Jel 72 (form II)........56
II.5 Purification an Crystallization of Jel 318..58
II.6 Discussion.................................61

III. **Structure Determination and Refinement**

III.1 Introduction..............................64
III.2 Determination and Refinement of Jel 72(I)...74

Chapter page
Data Collection.............................74
Analyses of Heavy Atom Derivatives...........77
Molecular Replacement.......................82
Structure Refinement........................93
Crystal Packing............................109
Structure Quality..........................115
Comparison of the Two Fab Fragments in the asymmetric unit............................118
The Antibody Combining Region..............126
III.3 Determination and Refinement of Jel 72(II)...........................131
Data Collection............................131
Molecular Replacement......................133
Structure Refinement.......................136
Crystal Packing............................137
Comparison with Form I.....................137
III.4 Determination and Refinement of Jel 318....138
Data Collection............................140
Molecular Replacement......................142
Structure Refinement.......................145
Crystal Packing............................154
Structure Quality..........................160
The Antibody Combining Region..............162
III.5 Refinement of Hed10.......................167
Structure Refinement.......................168

Chapter page

Structure Description......................176

III.6 Conclusions................................180

IV. Modelling of Antibody Combining Sites

IV.1 Introduction...............................183

IV.2 Model Building.............................188

IV.3 Comparison of Model and Crystal Structures.193

IV.4 Discussion.................................205

V. DNA Binding Models

V.1 Introduction................................210

V.2 Single Stranded DNA.........................214

V.3 Double Stranded DNA.........................220

V.4 Triple Stranded DNA.........................230

VI. Conclusions

VI.1 Discussion.................................238

VII. Bibliography................................249

VIII. Appendix...................................260

List of Tables

Table page

III-1 : Data Collection Statistics for Jel 72(I).......76

III-2 : Heavy Atom Coordinates and Occupancies.........79

III-3 : Refined Heavy Atom Parameters..................80

III-4 : Rotation Function Results for Jel 72(I)........86

III-5 : Phased Translation Function Results............87

III-6 : Refinement Statistics for Jel 72(I)...........101

III-7 : Crystal Contacts for Jel 72(I)................112

III-8 : RMS Deviations from Alignment for Jel 72(I)...122

III-9 : Data Collection Statistics for Jel 72(II).....132

III-10: Rotation and Translation Function Results for Jel 72(II)...............................134

III-11: Refinement Statistics for Jel 72(II)..........136

III-12: Data Collection Statistics for Jel 318........141

III-13: Molecular Replacement Results for Jel 318.....144

III-14: Refinement Statistics for Jel 318.............148

III-15: Crystal Contacts for Jel 318..................155

III-16: Unit Cells of Crystals Similar to Jel 318.....157

III-17: Refinement Statistics for Hed10...............171

IV-1 : Amino Acid Sequences for the Hypervariable Loops of nucleic acid binding Antibodies........189-190

IV-2 : RMS Deviations of Modelled Loops..............200

List of Figures

Figure page

I-1 : Structure of an IgG Antibody......................13

I-2 : The Immunoglobulin Fold..........................18

I-3 : Structure of an Fab Fragment......................20

II-1 : IEF Gel of Digests of Jel 72 and Jel 318...........52

II-2 : Purification of Jel 72 Fab Fragments...............53

II-3 : Crystals of Jel 72 Fab Fragments..................55

II-4 : Purification of Jel 318 Fab Fragments..............60

III-1 : Harker Construction for Phase Determination by Single Isomorphous Replacement....................68

III-2 : Harker Section (U,1/2,W) for Pt$(CNS)_6$ derivative...78

III-3 : Self Rotation Function for Jel 72(I)...............84

III-4 : Rotation and Phased Translation Searches........88-91

III-5 : Temperature Factor Distribution for Jel 72(I).102-103

III-6 : Regions of Good and Poor Density for Jel 72(I)....104

III-7 : Ramachandran Plots for Jel 72(I).............105-108

III-8 : Crystal Packing of Jel 72........................110

III-9 : Pseudo-4_1 Symmetry of Jel 72(I)...................113

III-10 : Luzatti Plot for Jel 72(I).......................117

III-11 : RMS Deviations from Alignment of Heavy and Light Chains of Jel 72(I).............................121

III-12 : Stereo Views of Aligned Loops................123-125

III-13 : Stereo View of Jel 72 Fv.........................127

III-14 : Stereo Views of Jel 72 CDRs......................129

III-15 : Ramachandran Plots for Jel 318...............149-150

Figure page

III-16 : Temperature Factor Distribution for Jel 318.......151

III-17 : Regions of Good and Poor Density for Jel 318......152

III-18 : Luzatti Plot for Jel 318..........................153

III-19 : Packing of Fabs in Different Crystals.............158

III-20 : Stereo View of Jel 318 Fv.........................163

III-21 : Stereo Views of Jel 318 CDRs......................165

III-22 : Ramachandran Plots for Hed10.................172-173

III-23 : Temperature Factor Distribution for Hed10.........174

III-24 : Luzatti Plot for Hed10............................175

III-25 : Stereo View of Hed10 Fv...........................178

III-26 : Stereo Views of Hed10 CDRs........................179

IV-1 : Hypervariable Loop Conformations.............194-199

IV-2 : Surface Potential for Crystal and Model Structures of Hed10............................209

V-1 : Electrostatic Surface Potential for Hed10.....216-217

V-2 : Electrostatic Surface Potential for Jel 72....222-223

V-3 : DNA Binding Model 1 for Jel 72 and poly(dG)•poly(dC)................................225

V-4 : DNA Binding Model 2 for Jel 72 and poly(dG)•poly(dC)................................227

V-5 : Electrostatic Surface Potential for Jel 318...233-234

V-6 : Model for Jel 318 Binding Triplex DNA............235

List of Symbols and Abbreviations

a,b,c	unit cell axes
B	thermal motion parameter
C_L	constant light chain domain of an Fab fragment
C_H	constant heavy chain domain of an Fab fragment
d_{min}	minimum interplanar spacing of diffraction data
F	structure factor
F_{obs}	observed structure factor
F_{calc}	calculated structure factor
F_H	heavy atom structure factor
F_P	protein structure factor
F_{PH}	protein plus heavy atom structure factor
F_{PH+}, F_{PH-}	Friedel pair of derivative data
$\|F\|$	structure factor amplitude
$\|F_{obs}\|$	observed structure factor amplitude
Fab	antigen binding fragment of an antibody
Fc	constant (crystallizable) fragment of an antibody
Fv	fragement containing the varaible domains of an antibody
f_m	figure of merit
I	intensity
IgA	Immunoglobulin A
IgG	Immunoglobulin G
IgM	Immunoglobulin M
MIR	multiple isomorphous replacement

NMR	nuclear magnetic resonance
R	standard crystallographic residual
RMS	root-mean-square
V_L	variable light chain domain of an Fab fragment
V_H	variable heavy chain domain of an Fab fragment
α_c	calculated structure factor phase
α_m	observed (from MIR) structure factor phase

I.1 Anti-DNA Antibodies

Antibodies are proteins that are produced in vertebrates by the immune system in response to a challenge by foreign macromolecules. They are found in serum and on the surface of B lymphocyte cells. The role of antibodies in the immune system is to bind to the foreign molecules and mark them for elimination from the blood stream.

The primary role of the immune system is recognition. It must recognize the presence of foreign macromolecules and be able to discriminate between these molecules and the molecules that are natural components of the animal. The repression of reaction by the immune system with self molecules is referred to as tolerance. When there is a breakdown in tolerance the immune system produces antibodies that are specific for endogenous macromolecules. This can result in a number of autoimmune disorders ranging from arthritis to systemic lupus erythematosus (SLE). When the self molecule is a major constituent of normal serum, antibody-antigen complexes circulating in the blood can reach abnormally high levels. A common clinical manifestation of these autoimmune disorders is glomerulonephritis, resulting from the deposition of antibody-antigen complexes in the glomeruli of the kidney. These conditions can lead to renal failure and may even be fatal.

DNA is a natural component of serum. It is introduced into the blood stream as a result of cell damage, but is normally cleared from the system. In the human autoimmune disorder SLE, and a very similar condition in mice, antibodies are produced that bind to DNA. Other antibodies that bind to proteins normally found in cell nuclei are also observed in the sera of patients suffering from SLE, as well as antibodies that bind phospholipids (Stollar, 1986). DNA binding antibodies with a wide range of affinities for both native and denatured DNA can be isolated from the sera of autoimmune mice. Moreover, DNA isolated from circulating immune complexes has been found to be enriched in guanine and cytosine (Sano and Morimoto, 1982). This observation could indicate that either antibodies interact preferentially with guanine and cytosine containing regions of DNA, or that these regions of DNA are less sensitive to digestion by nucleases present in the blood stream.

There is some controversy as to whether the production of autoimmune anti-DNA antibodies is actually elicited by nucleic acids or by some other macromolecule. Animals injected with bacterial lipopolysaccharide produce antibodies that bind DNA (Fournie et al. 1974; Dziarski, 1982). Bacterial lipopolysaccharide is a polyclonal activator that can stimulate antibody production independent of antigen. Thus, there is some background level of DNA binding reactivity encoded in the germ line genes of antibody molecules.

Moreover, anti-DNA antibodies isolated from the sera of SLE patients can be inhibited by cardiolipin (Koike et al. 1982), while antibodies that bind cardiolipin but do not bind DNA can also be detected in autoimmune sera (Harris et al., 1983). However, studies on several clonally related anti-DNA autoantibodies derived from a single autoimmune mouse (Shlomchik et al., 1987) found that numerous somatic mutations had occurred in the variable region genes, suggesting that these antibodies had been positively selected for by DNA.

Regardless of their origin and cross-reactivity with other biological macromolecules, antibodies that do bind duplex DNA have been directly implicated in the pathogenesis of SLE (Lange, 1978; Abu et al., 1981). Thus, one focus of these studies is to determine the structural features of anti-DNA antibody combining regions to further elucidate how they might recognize DNA.

DNA binding antibodies can also have potential uses as diagnostic and biochemical reagents. Antibodies specific for modified bases can be used to detect the presence of carcinogen-DNA adducts (Poirer, 1981; Munns and Liszewski, 1980; Strickland and Boyle, 1984) which has implications for the diagnosis and treatment of cancer. Antibodies that bind 'A' helical double stranded RNA (dsRNA) can be used to detect the presence of dsRNA in virus infected cells (Stollar and Stollar, 1970; Ng et al., 1983). A monoclonal antibody specific for triple stranded DNA has been shown to be able to detect the presence of triplex in nuclei and in chromatin, suggesting that triple stranded nucleic acids may have a biological role (Lee et al., 1987). With knowledge of the structures of a few anti-DNA antibodies and a working model for how they might recognize DNA, the possibility exists that other antibodies could be engineered to detect specific conformations or sequences of DNA.

Another aspect of this work is to study the combining regions of anti-DNA antibodies to gain insights into the general features of protein-nucleic acid interaction. Questions arise on how antibodies recognize nucleic acids and how is this interaction different from, or similar to, other proteins that bind DNA. Fortunately, there is another source of anti-DNA antibodies other than from autoimmune sera. DNA binding antibodies can be induced by immunizing animals with synthetic polynucleotides. DNAs that can induce antibody production are usually repeating polymers, suggesting that the antibodies are recognizing some structural feature or conformation adopted by the DNA. Antibodies have been raised that bind to repeating polymers of 'A' , 'B' or 'Z' form DNA (Anderson et al., 1988). In this way antibodies specific for 'Z' DNA have been obtained by immunizing animals with brominated poly(dGC) (Stollar, 1986). The 'Z' DNA binding antibodies are able to discriminate between the 'Z' form of poly(dGC) and native DNA. They also do not bind single stranded DNA, single or double stranded RNA, RNA-DNA hybrids, or the synthetic polynucleotides poly(dG), poly(dC), poly(dAT), poly(dA)·poly(dT), or brominated poly(dG)·poly(dC). The advantages of studying monoclonal antibodies of this type stem from the fact that they bind to DNA with a specific base sequence, rather than the duplex specific autoimmune antibodies which generally recognize 'B' form DNA with little or no base sequence specificity. Therefore, we have undertaken the study of the structural basis for antibody-DNA recognition using x-ray crystallographic techniques and antibodies that bind DNA polymers with specific base sequences.

Three particular monoclonal antibodies that we have studied are Hed10, Jel 72 and Jel 318. Hed10 is a murine autoimmune antibody that binds single stranded DNA with a marked base preference for polynucleotides containing thymine. Hed10 binds poly(dT) with a binding constant that is an order of magnitude higher (10^{-7} M^{-1} versus 10^{-6} M^{-1}) than for either poly(dbrC) or poly(dU). Hed10 does not bind poly(dC), poly(rU), poly(dA), poly(dA)·poly(dT), or either native or heat denatured calf thymus DNA, and also does not bind cardiolipin (Lee et al., 1982). Jel 72 is a mouse monoclonal antibody that binds the right handed duplex formed by poly(dG)·poly(dC). The binding of Jel 72 to various polynucleotides has been examined by solid phase radioimmunoassay and it was found to recognize this sequence and to not bind to several related DNA polymers, including: poly(dG)·poly(dm^{5}C), poly(dGC)·poly(dGC), poly[d(Gm^{5}C)]·poly[d(Gm^{5}C)], poly[d(TTC)]·poly[d(GGA)], poly(dI)·poly(dC), poly(dI)·poly(dm^{5}C), poly(dI)·poly(rC), poly(rG)·poly(dC), poly(rI)·poly(dm^{5}C), and to also not bind to the single stranded polymers poly(dG), poly(dC), and poly(dm^{5}C) (Lee et al., 1984).

Jel 318 is also a mouse monoclonal antibody, however, it is specific for triple stranded DNA of the sequence poly(dTm^{5}C)·poly(dGA)·poly(dTm^{5}C$^+$) (Lee et al., 1987). Solid phase radioimmunoassay results show that Jel 318 does bind, albeit less strongly, to DNA polymers that can disproportionate to form triplex, but this conversion may, in fact, be induced by the antibody. These polymers include: poly[d(Tm^{5}C)]·poly[d(GA)], poly[d(TC)]·poly[d(GA)], poly[d(TTm^{5}C)]·poly[d(GAA)], and poly(dA)·poly(dT). Jel 318 does not bind to poly[d(TC)]·poly[d(Gm^{6}A)], which can not form triplex because the methyl group at the 6 position of adenine blocks Hoogsteen base pairing. Jel 318 also does not bind to either native or heat denatured calf thymus DNA, nor to the double stranded DNA polymers poly(dG)·poly(dm^{5}C), poly[d(Tm^{5}Cm^{5}C)]·poly[d(GGA)], poly[d(TG)]·poly[d(CA)], or poly[d(Gm^{5}C)·poly[d(Gm^{5}C)] (Lee et al., 1987).

The three dimensional crystal structures of the antigen binding (Fab) fragments from the antibodies Hed10, Jel 72 and Jel 318 have been determined in the laboratory of Dr. Wayne F. Anderson. The research performed on this project contains significant contributions from several individuals. The purification and crystallization of the Fab fragments from Hed10 and Jel 318 was performed by Mr. Amechand Boodhoo. The determination of the structure and crystallographic refinement of Hed10 was achieved by Dr. Mirek Cygler. Dr. Cygler also determined the rotation and translation parameters for the molecular replacement solution of Jel 318, while an initial refinement of this structure was performed by Dr. Alastair K.S. Muir. Dr. Muir also assisted in the solution of the first crystal form of Jel 72 by correctly interpreting the rotation search results. My contribution to the project involved the purification and crystallization of two separate crystal forms of the Fab fragments from Jel 72. I also applied Dr. Muir's results to the translation search to complete the molecular replacement solution of the first crystal form of Jel 72. The structure determination of the second crystal form of Jel 72 was achieved entirely by myself. I also performed the crystallographic refinements of both crystal forms of Jel 72 and further refined the structures of Hed10 and Jel 318.

This project was undertaken with the collaboration of Dr. Jeremy S. Lee in the Department of Biochemistry at the University of Saskatchewan in Saskatoon, Saskatchewan. The production of immunoglobulins and the characterization of their nucleic acid binding specificity was performed in Dr. Lee's laboratory. The determination of the amino acid sequences for the antigen binding regions of the antibodies was performed by Dr. M. Michelle Berry working under the supervision of Dr. Lee. When I obtained this amino acid sequence information, I derived models for the antibody combining regions and compared them to the structures that we had determined using X-ray crystallographic techniques. These results proved encouraging and I, therefore, endeavoured to derive models for the combining regions of other nucleic acid binding antibodies for which no structural information is available. I have also proposed preliminary, working models for the interaction of these antibodies with their respective nucleic acid antigens and propose experiments to test these models and further elucidate the nature of antibody-nucleic acid interactions.

In the sections to follow there will be an introduction to antibody structure, and a survey of the structural studies that have been conducted on antibody-antigen complexes and on DNA binding proteins complexed with DNA. More complete descriptions of the antibody-antigen complexes and of the structural studies of protein-nucleic acid interactions may be found in reviews by Davies et al.(1990) and Steitz (1990), respectively.

I.2 Structure of Immunoglobulin G

Antibodies are the prototype for the immunoglobulin class of protein structure. They are multimeric and multidomain proteins consisting of two Heavy (H) and two Light (L) polypeptide chains. Antibodies are large proteins with an approximate M_r of 150,000. The L chain is about 220 amino acids long with an M_r of 25,000, while the length of the H chain varies between 450 to nearly 600 amino acids depending on the class to which the antibody belongs. The M_r for H chains that belong to the Immunoglobulin G (IgG) class is approximately 50,000.

Although there are five major classes of antibody molecules, characterized by the primary structure of their H chains and by their biologic function, antibodies that belong to the IgG class are by far the most prevalent in normal serum, the most widely studied and the best characterized. Light chains are of two types, either kappa or lambda. Light chains belonging to the kappa class are more common in antibodies derived from mice, with about 95% of antibodies possessing light chains of this type. While in humans the proportion of antibodies possessing kappa light chains is approximately 50%. As with the Heavy chains, the two different classes of Light chains differ in their primary structure such that they can be distinguished antigenically. While the various classes of H and L chains differ in their amino acid sequences, they all possess a similar domain structure, the immunoglobulin fold, that is characterized by an internal disulphide loop that encompasses approximately 50 to 70 amino acid residues. All of the antibodies that are the subject of this thesis belong to the IgG class of heavy chains and have kappa light chains. Therefore, the description of antibody structure and their biologic roles will concentrate on antibodies that belong to the IgG class.

Antibodies are produced by B lymphocyte cells which arise by differentiation of the bone marrow stem cells. Through a sequential process of gene rearrangement, the B cell expresses a series of genes that encode the complete Heavy and Light polypeptide chains of the antibody molecule. The first step in B cell maturation involves the rearrangement of the heavy chain genes resulting in a cell referred to as a pre-B lymphocyte. The rearrangement of the kappa light chain genes is the next step in B cell differentiation. If the rearrangement of the kappa genes is nonproductive and does not produce a viable kappa light chain polypeptide, the B cell rearranges the lambda light chain genes and, if a productive rearrangement occurs, produces light chain polypeptides of type lambda. The L and H chain polypeptides can then combine to form an intact immunoglobulin molecule. This immunoglobulin belongs to the IgM class of antibodies, remaining bound to the membrane at the surface of the cell, which is now referred to as a virgin B lymphocyte. The next step in B cell differentiation is triggered by an encounter with an antigen for which the expressed immunoglobulin has binding specificity.

The maturation of the immune response is characterized by a heavy chain switch. After a productive encounter with antigen the B cell becomes activated and undergoes further gene rearrangements to produce antibodies with Heavy chains that belong to the IgG class. The B cell no longer produces membrane bound IgM, but instead produces IgG antibodies that are secreted into the blood system. The antibody produced is a dimer composed of two covalently linked heavy chain polypeptides each interacting with a light chain protein.

The light chain folds into two globular domains termed V_L and C_L, while heavy chains belonging to the IgG class consist of four domains designated V_H, C_{H1}, C_{H2} and C_{H3}. Each domain is about 110 amino acids long and folds into a basic structural motif termed the immunoglobulin fold. These domains associate with other domains to form discrete portions of the antibody molecule. The V_L and V_H domains combine to form the Fv. The fragment containing the Fv and the C_L and C_{H1} domains is referred to as the Fab. The Fv portion of an Fab fragment is also called the variable region, while the portion of the Fab consisting of the C_L and C_{H1} domains is referred to as the constant region. Within an Fab fragment the C_{H1} domain may also be referred to simply as C_H. The C_{H2} and C_{H3} domains of both heavy chains combine to form the Fc portion of the antibody molecule. Antibodies that belong to the IgG class have attached carbohydrate that is covalently linked through the side chain of an asparagine residue in the C_{H2} domain of the heavychains.

The polypeptide regions linking these domains are flexible. The region between the variable and constant domain pairs of the Fab fragment is called the elbow region, while the polypeptide between C_{H1} and C_{H2}, which links the Fab to the Fc, is referred to as the hinge region. The hinge region is susceptible to cleavage by proteolytic enzymes such as papain, pepsin and trypsin. Incubation of intact antibody molecules with papain is routinely used for the preparation of free Fab fragments. Digestion of antibodies with pepsin results in an Fab'_2 fragment, in which two Fab fragments are covalently linked through a disulphide bond (Figure I-1).

A landmark study of a large number of the amino acid sequences of antibodies (Wu and Kabat, 1970) revealed that, within the class of H and L chains, the amino acid sequence is conserved for the C_{H1}, C_{H2}, C_{H3} and C_L domains. The amino acid sequences of the V_L and V_H domains contained regions of sequence variability which were unique for any particular antibody. It is within these latter domains that the antibody combining region is located. Those portions of the amino acid sequence that are conserved within the V_H and V_L domains are designated framework (FW) regions. The remaining regions possess considerable variability both in their length and amino acid sequences.

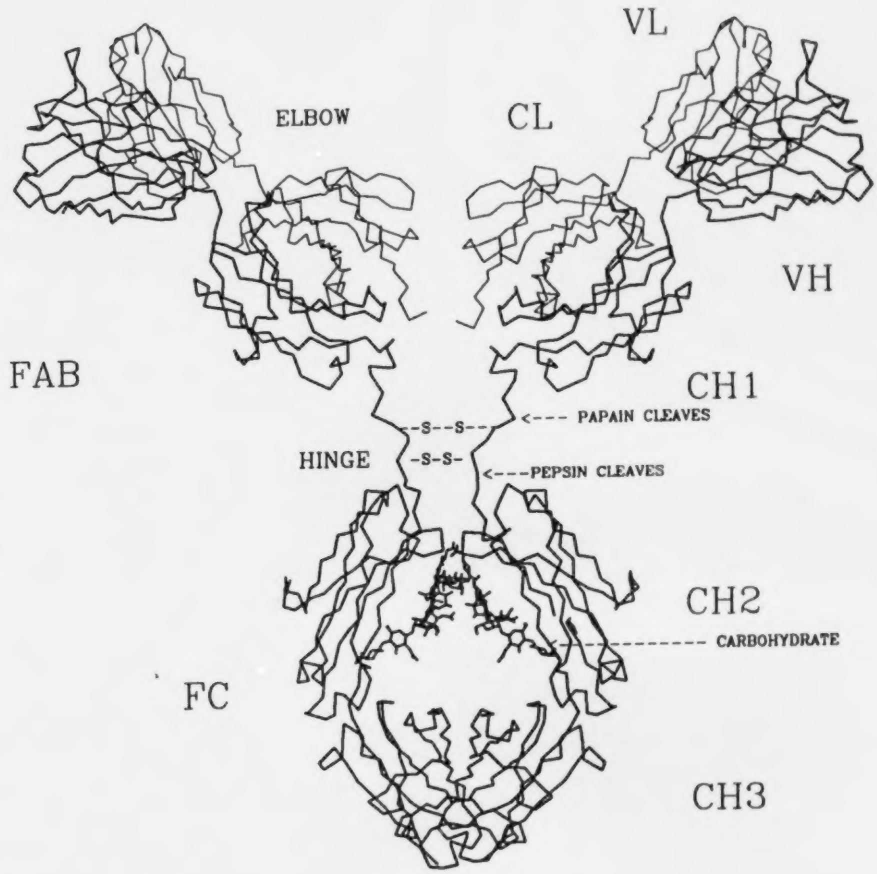

VL
ELBOW
CL
VH
FAB
CH1
-S--S-
<--- PAPAIN CLEAVES
HINGE
-S-S-
<---PEPSIN CLEAVES
CH2
CARBOHYDRATE
FC
CH3

Figure I-1: The structure of an Immunoblogulin G antibody. The Heavy chains (thick lines) fold into four discrete domains (C_{H1}, C_{H2}, C_{H3} and V_H) and the Light chains (thin lines) into two domains (C_L and V_L). The model was constructed form the Fc fragment (Deisenhofer, 1981) and the Fab fragment Kol (Matsushima et al., 1977), and a short segment of polypeptide for the hinge region. The positions where the antibody is susceptible to cleavage by papain and pepsin to produce free Fab and Fab'$_2$ fragments, respectively, are also indicated.

The areas of variable length and amino acid sequence in the V_H and V_L domains are termed complementarity determining regions (CDR), and are involved in the recognition and binding of antigen. The antibody combining site is comprised of six CDRs, three each from V_H and V_L. The variation in the amino acid sequences of these CDRs arises from three processes: 1) multiple genes, 2) somatic mutation, and 3) somatic recombination (Honjo, 1983). The amino acid sequence of the V_L domain is encoded by two separate genetic elements, the V gene and the J (joining) gene segment (Weigert et al., 1980). There are multiple copies of each gene. Somatic recombination between one copy of each gene results in a gene that encodes the amino acid sequence for a complete V_L domain. The gene that encodes the V_H domain is generated by recombination between three separate genes, the V, D (diversity), and J gene segments (Davis et al., 1980; Early et al., 1980). The junctions between these gene segments are not precisely defined with the result that all three reading frames of the D and J genes can be used to generate further sequence variability. Somatic mutation of the assembled V_H and V_L genes can also contribute to the generation of amino acid sequence variability of the antibody combining site.

Antibody populations in normal serum are usually heterogeneous, consisting of a vast array of antibodies with diverse binding specificities. However, when there is some immune system dysfunction plasmacytomas or myelomas may result. Myelomas arise spontaneously in humans and produce significant amounts of a single homogeneous antibody. The initial structural studies on antibodies and their proteolytic fragments were restricted to these myeloma proteins since sufficient quantities of pure antibody could not otherwise be obtained. With the advent of hybridoma technology (Koehler and Milstein, 1975), monoclonal antibodies with a single defined specificity can be obtained in large amounts. The genes for these antibodies can be cloned and sequenced and their amino acid sequences deduced in a straightforward manner. This has resulted in a vast increase in the amount of amino acid sequence information pertaining to antibody combining regions. Any method that can place this sequence information within a structural context will be widely applicable.

Most of the information that has been obtained on antibody structure is the result of X-ray crystallographic analyses of antibodies and their proteolytic fragments. These studies established the basic structural fold of the immunoglobulin molecule and explained the observations derived from the sequencing studies. The principal secondary structural motif of the immunoglobulin fold is two anti-parallel β-pleated sheets. These sheets possess the characteristic right handed twist observed in β-sheet structures and stack in an aligned fashion, with an angle of approximately 30° between the strands of the stacked sheets (Chothia and Janin, 1981). This type of association is observed in the V_L, V_H, C_L and C_H domains of immunoglobulins. The sheet-sheet interactions that govern the association of these domains to form the constant and variable regions of the immunoglobulin molecule, however, do not conform to an aligned packing mode. The association of the C_L and C_{H1} domains and the dimerization of the heavy chain C_{H3} domains demonstrates a second stereochemical theme for sheet-sheet interactions in which the β-sheets pack in an orthogonal orientation with an angle between the sheets of approximately 90° (Chothia and Janin, 1982). The β-sheet interactions observed in the association of the V_H and V_L domains to form the variable region of the antibody, however, do not necessarily conform to either the aligned or orthogonal orientations (Chothia et al, 1985). The interface between the V_L and V_H domains is influenced by the lengths and amino acid sequences of the antibody complementarity determining regions which can contribute approximately 25% of the residues that pack within the domain interface. The influence that these hypervariable residues have on the sheet-sheet interactions is such that the angle between the associated sheets is often observed to be midway between that of the aligned and orthogonal classes with a value of approximately 50°.

The region between the two stacked β-sheets is tightly packed with the side chains of hydrophobic amino acids. There is an intrachain disulphide bridge that covalently links the two β-sheets and there can also be an interchain disulphide bridge linking the C_L domain to the C_{H1} domain. The constant domains (C_L, C_{H1}, C_{H2}, C_{H3}) contain one three-stranded and one four-stranded β-sheet, while the variable domains (V_L and V_H) possess an extra two β-strands resulting in the formation of one four-stranded and one five-stranded β-sheet (Figure I-2). The C_L and C_{H1} domains interact extensively via their four-stranded β-sheets to form the constant region of an Fab fragment. The V_L and V_H domains associate through their five-stranded β-sheets to form the variable region of the molecule. These associated domains are related by a rotation of approximately 180° that is referred to as the pseudodiad angle. Typical values for the pseudodiad are 165° - 170° for the constant region and 170° - 180° for the variable region.

The antibody combining region is located at the end of the V_H and V_L domains that is furthest from the constant domain pair. The three-dimensional shape and chemical nature of the combining region is largely determined by the lengths and amino acid sequences of six polypeptide segments that link the strands of the β-sheets and form external loops at this end of the variable domain pair. The portions of the variable domains which form the conserved β-sheet structure correspond to the framework regions of the amino acid sequence.

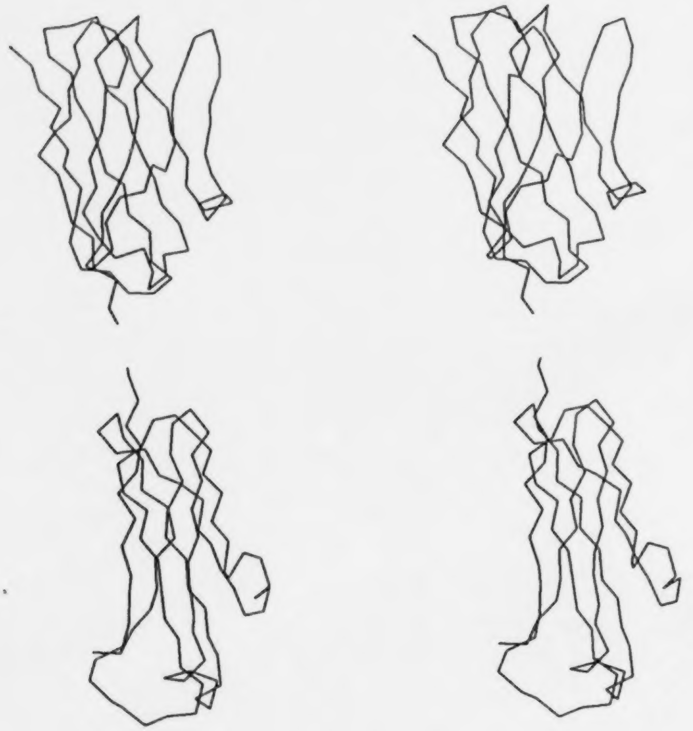


Figure I-2: The Immunoglobulin Fold. Stereo representation of the immunoglobulin fold as observed in the V_H domain (top) and the C_{H1} domain (bottom) of the Fab fragment Hed10 (Cygler et al., 1987).

The segments of the variable domains that form the external loops correspond roughly to the complementarity determining regions. These regions are also called hypervariable loops, and are designated H1, H2, and H3 or L1, L2 and L3, corresponding to the H and L chains and approximately equal to the CDRs in position. The structure of an Fab fragment is illustrated in Figure I-3.

The amino acid sequences for the Fv domains of the six nucleic acid binding antibodies that are the subject of this thesis are listed in Appendix A. Also listed in Appendix A is the correlation between the CDRs, as determined from the amino acid sequences, and the designation of the hypervariable loops derived from an analysis of known Fab fragment structures. The numbering scheme commonly employed when referring to antibody combining regions is that suggested by Kabat et al. (1987). This numbering scheme is also described in Appendix A.

A survey of known Fab fragment structures has revealed that some of the hypervariable loops belong to discrete conformational classes, or canonical structures, while others display greater conformational variability (Chothia and Lesk, 1989). The H3 loop, in particular, has proven to be capable of a great amount of structural diversity. With the possible exception of this loop, many of the other hypervariable loops in antibodies of unknown structure may be in conformations similar to those in previously determined structures.

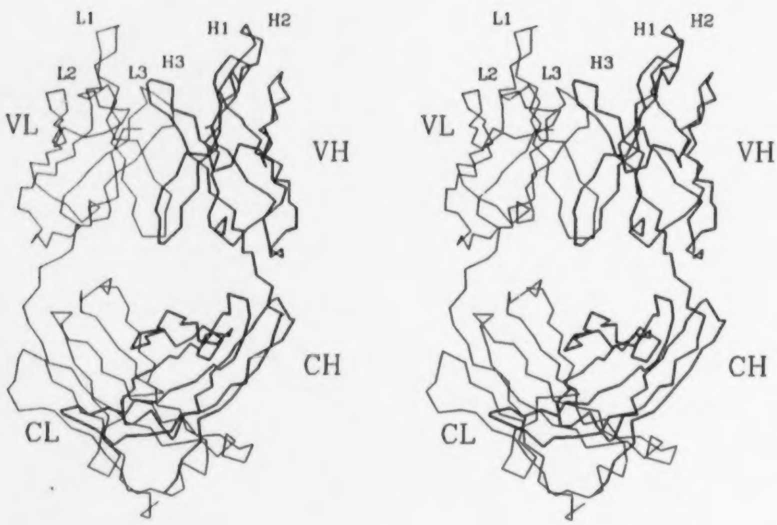

L1
H1
H2
L2
L3
H3
VL
VH
CH
CL
L1
H1
H2
L2
L3
H3
VL
VH
CH
CL

Figure I-3: The Structure of an Fab Fragment. The domain structure of Fab fragment Hed10 is illustrated. The C_{H1} domain of the Heavy chain (thick lines) combines with the C_L domain of the Light chain (thin lines) to form the constant region of the molecule, while the V_H and V_L domains combine to form the variable region. The antibody combining region is located at the top of the figure and the locations of the antibody CDRs are indicated.

The importance of the immunoglobulin fold in biology is underscored by the fact that it has been found to occur in many other proteins. The functions of proteins that contain at least one immunoglobulin like domain are not restricted to the immune response, but seem to play a general role in molecular recognition at cell surfaces. Some of these proteins that are not involved in immune recognition include the neural cell adhesion molecule, NCAM (Cunningham et al. 1987), and the family of receptor molecules related to the platelet derived growth factor receptor, PDGF (Yarden et al., 1986). The superfamily of immunoglobulin related proteins also includes the T-cell receptor proteins (Hood et al., 1985) as well as subunits of the Class I and Class II Major Histocompatibility Complexes, MHC (Davis and Bjorkman, 1988). The existence of the immunoglobulin fold in many of these proteins has been postulated to occur based on amino acid sequence similarity with immunoglobulin domains of known structure (Williams and Barclay, 1988).

From an analysis of the structures of immunoglobulin domains, Chothia and colleagues (1988) have proposed that there are forty sites critical to the conserved framework structure of antibody variable domains. The known amino acid sequences of the V_α and V_β regions of the T-cell $\alpha\beta$ receptor contain identical, or very similar, residues at these positions and are, thus, proposed to possess a framework structure very similar to antibody domains. Chothia and coworkers also show that the V_α - V_β sequences contain regions of hypervariability. These regions are in different conformations than the hypervariable regions that are observed in antibodies, and some of these regions are proposed to interact with the MHC proteins.

The amino acid residues that are considered crucial to the establishment of the immunoglobulin fold include those residues of the conserved β-strands that form the hydrophobic core of the domain and stabilize the interaction of the β-sheets. The conserved intrachain disulphide bond is often considered to be a defining characteristic of immunoglobulin like domains, although a functional antibody has been obtained that lacks a variable region disulphide bridge (Rudikoff and Pumphrey, 1986).

The central role that the immunoglobulin fold appears to play in cellular recognition processes may arise from the fact that it is able to accommodate different lengths of polypeptides, with variable amino acid sequences, in between the two cysteines of the conserved disulphide. This allows for the basic structural fold of the molecule to remain intact while the surface displayed by these variable loops can be altered to optimize interactions with the macromolecules involved in specific recognition processes. A common feature of all of the proteins that are proposed to contain immunoglobulin like domains is that they are extracellular. Thus, it has been suggested that the immunoglobulin fold may be conserved in biology partly because it is resistant to the action of proteases that are present in the extracellular environment (Williams, 1987).

I.3 Antibody-Antigen Complexes

A number of antibody Fab fragment structures have been determined in complex with specific antigens. These range from complexes with small molecular weight haptens to Fab fragments bound to complete proteins. The salient features derived from these studies may not be directly applicable to antibodies that bind nucleic acids. Nevertheless, insights derived from these works are of a sufficiently general nature that they merit consideration as possible mechanisms employed by DNA binding antibodies.

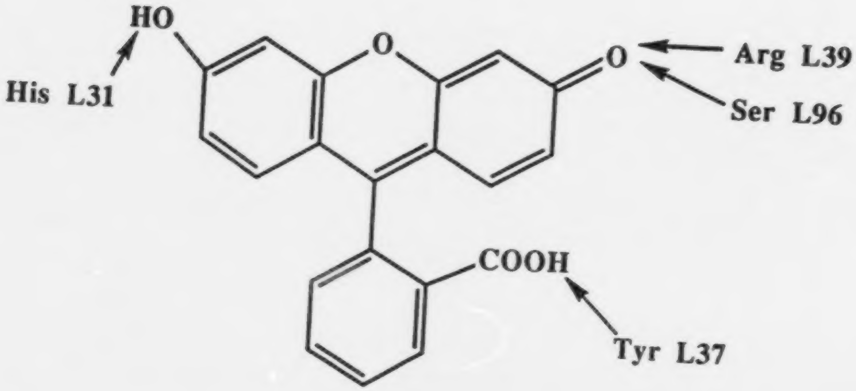
McPC603 and phosphocholine: The myeloma protein McPC603 was one of the first Fab fragments to have its structure determined by X-ray crystallography (Segal et al., 1974). Its structure has been determined both alone (Satow et al. 1986) and in complex with phosphocholine (Davies et al. 1990), for which it has a binding constant of $0.2 \times 10^6 M^{-1}$ (Metzger and Otchin, 1971). The choline moiety binds in a pocket within the combining site while the phosphate is on the surface. The positive charge on the choline is neutralized by interaction with aspartic acid L91[1] and to a lesser extent with glutamic acid H35. The negatively charged phosphate can form a hydrogen bond with the NH1 and OH groups, respectively, of arginine H52 and tyrosine H33. Charge neutralization of the phosphate is also accomplished by the positive charge on lysine H54. These studies helped to establish McPC603 as perhaps the prototypical Fab fragment. Modern biochemical and immunological experiments continue to refer to this pioneering work.

---

[1] The numbering scheme used when referring to antibody combining regions is that suggested by E.A. Kabat and co-workers (1987). The number is prefixed by either an H or an L to indicate whether the residue is from the Heavy or Light chain, respectively.

4-4-20 and Fluorescein: The crystal structure of the Fab fragment 4-4-20 in complex with fluorescein has been determined at 2.7 Å resolution (Herron et al., 1989). This particular monoclonal antibody has a binding constant for fluorescein of 3.4×10^{10} M^{-1}. The crystals were grown from a 47% solution of 2-methyl-2,4-pentanediol (MPD) which has the effect of lowering the affinity for fluorescein 300-fold. This structure is of interest because it crystallizes in a space group nearly isomorphous with the autoimmune antibody Fab fragment BV04-01, which has a specificity for single stranded DNA. The crystal structure of BV04-01 in an uncomplexed form had been previously determined (Herron et al., 1987). BV04-01 and 4-4-20 share nearly identical light chains, differing in only six residues. The L chains are in similar conformations and are involved in a majority of the crystal packing interactions resulting in the two Fab fragments crystallizing in very similar space groups.

The fluorescein binds in a slot formed by the side chains of the aromatic residues tryptophan L101, tryptophan H33 and tyrosine L37. The two enolic groups on the xanthonyl ring are within hydrogen bonding distance of histidine L31 and arginine L39, respectively. The guanido group of arginine L39 is close enough for there to be an electrostatic interaction with the enolic oxygen. There is also the potential for a hydrogen bond between this oxygen and the side chain of serine L96. An important point is that there is no formal charge neutralization between the Fab fragment and the negatively charged carboxylic acid moiety on the hapten. The carboxylic acid group is exposed to solvent. A hydrogen bond between the hydroxyl of tyrosine L37 and an oxygen atom of the carboxylic acid partially neutralizes the negative charge at this end of the fluorescein molecule.



Fluorescein

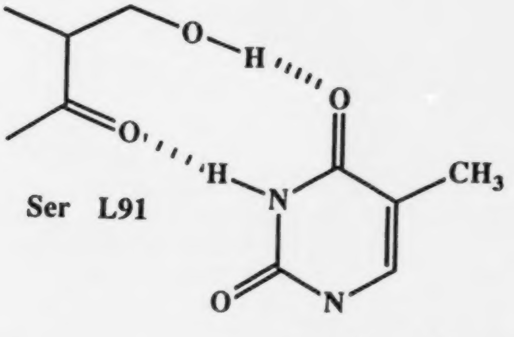
These two structures are of particular relevance to DNA binding antibodies in the specific interactions between the protein and the negatively charged phosphate and carboxylic acid groups. DNA contains many negatively charged phosphodiester groups. The molecular mechanisms by which antibodies such as McPC603 interact with phosphate groups establish principles that may be applicable to the recognition of DNA by DNA binding antibodies. Formal neutralization of the negative charge on the sugar-phosphate backbone of DNA may be an important aspect, but it is not necessarily a requirement for binding. The potential contributions from hydrogen bonding with neutral polar residues as well as the protein backbone should also be considered. In fact, formal neutralization of the negative charge on the DNA backbone is seen to play a relatively minor role in the recently determined crystal structure of a trinucleotide-Fab complex (Herron et al., 1991).

BV04-01 and d(pT)$_3$: BV04-01 is a murine autoimmune anti-DNA antibody with a specificity for single stranded DNA (ssDNA) and a base preference for thymine. The binding of this antibody to 5'-phosphorylated oligodeoxythymidylic acids is relatively insensitive to ionic strength and can not be detected for ssDNA that is less than 6 nucleotides in length (Ballard and Voss, 1985).The affinity of BV04-01 for single stranded poly(dT) is similar, but not identical, to that exhibited by Hed10. While Hed10 is also a murine autoimmune antibody that binds ssDNA containing thymine, the dependence of this binding on ionic strength has been interpreted to indicate that Hed10 interacts with two phosphate groups on the DNA and that its active site covers four bases (Lee et al., 1982). The interactions that are observed in the crystal structure of the BV04-01 Fab-trinucleotide complex may be particularly relevant to our understanding of the binding of single stranded poly(dT) by Hed10.

The structure of the BV04-01 Fab-trinucleotide complex has been determined at 2.66 Å resolution and refined to an R-value of 0.191 (Herron et al., 1991). The trinucleotide in this complex is in an irregular conformation that is complementary to the surface of the antibody combining site. The nucleotide at the 5' end interacts with the Heavy chain, the central nucleotide interacts with residues of both chains, while the nucleotide at the 3' end makes contacts with residues of the Light chain. Two oxygen atoms of the 5' phosphate group of the first nucleotide are in positions to form hydrogen bonds with the side chain hydroxyl group and amide group of serine H52a and asparagine H53, respectively. The deoxyribose ring and thymine base of this nucleotide contact the protein through the side chain and backbone atoms of tryptophan H100a. The O4 atom of this thymine is within hydrogen bonding distance (2.8 Å) of the peptide nitrogen atom of glycine H98. All of the protein contacts with the thymine base of this nucleotide are mediated by residues in the third

CDR of the heavy chain.

In contrast, the thymine base of the central nucleotide makes contacts with both the H and L chains. The base intercalates between the aromatic side chains of tyrosine L32 and tryptophan H100a, and forms two hydrogen bonds with serine L91. The N3 atom of this thymine is 2.7 Å away from the backbone carbonyl oxygen of serine L91 while the O4 atom is a distance of 2.9 Å from the side chain hydroxyl of this serine.



Thymine-2

The authors suggest that this double hydrogen bonding pattern may be responsible for the immunodominance of thymine and guanine in many of the autoantigens observed in SLE (Munns et al., 1984). A guanine base could also be capable of forming the observed double hydrogen bonds through its N1 and O6 atoms. However, guanines will not fit within the BV04-01 combining site if they are placed onto the sugar-phosphate backbone of the trinucleotide in the complex structure.

Additionally, binding of the first two nucleotides seemed to require that the thymine bases be in the **anti** configuration, while single stranded deoxynucleotides containing guanine prefer the **syn** orientation (Saenger, 1984).

The phosphate group of the central nucleotide is involved in a charge-charge interaction with the side chain of arginine H52, but no other ion pair is observed in the complex. Histidines L27d and L93 contact the deoxyribose portion of the third nucleotide and a potential hydrogen bond exists between the imidazolium NH of histidine L93 and the O2 atom of the thymine base. No specific interactions with the phosphate group of the third nucleotide are observed.

The structure of the Fab fragment from BV04-01 in an uncomplexed form has also been determined (Herron et al., 1991) and refined to an R-value of 0.246 for data to 2.0 Å. By comparing this structure to the structure of the Fab fragment in complex with the trinucleotide, the question of whether there is an "induced fit" of the antibody combining site to the ligand can be addressed. The authors conclude that a significant rearrangement of the combining site occurs when the Fab binds the trinucleotide.

The quaternary structure of the Fab is slightly different in the two crystal forms. The elbow angle that relates the variable region to the constant region is 172° in the free Fab and 175° for the Fab in the complex. The association of the V_H and V_L domains, as defined by the pseudodiad angle, is also different. The pseudodiad of the variable region in the complex structure is 179° while it is 174° in the free Fab. The RMS deviations between pairs of Cα atoms after the superposition of the individual domains is 0.45 Å for V_L and 0.54 Å for V_H. By comparing only those Cα atoms that were determined to be in regions with similar tertiary structures in the two Fabs, the differences in the atomic positions that could be attributed to this readjustment of the variable domains could be determined. An RMS displacement of 1.45 Å was found to have occurred with 62% of the Cα atoms exhibiting a displacement of greater than 1.0 Å from their positions in the unliganded Fab structure. A similar analysis of the differences in the association of the C_L and C_{H1} domains found that an RMS displacement of only 0.65 Å had occurred with fewer than 10% of the Cα atoms exhibiting a displacement of greater than 1.0 Å.

These results are in agreement with previous proposals (Colman et al., 1987; Colman, 1988; Colman et al., 1989) that suggested that the complementarity between the antibody combining site and the antigen can be improved by adjustments in the relative orientations of the V_H and V_L domains during antigen binding. While this adjustment in the quaternary association of the variable domains may play a significant role in the formation of a stable antibody-DNA complex, a far greater conformational change is observed in the tertiary structures of two of the hypervariable loops. Herron et al.,

(1991) attribute the alteration of the conformations of the L1 and H3 hypervariable loops to an "induced fit" of the combining site to the trinucleotide. Residues in the H3 loop interact with the first two nucleotides and contacts between the third nucleotide and the L1 hypervariable loop are observed in the complex. An RMS displacement from the atomic positions observed in the unliganded Fab structure of 0.80 to 3.33 Å for the amino acids of the L1 loop and from 1.44 to 2.99 Å for the residues in the H3 loop is concluded to have occurred. Thus, extensive adjustments of the antibody combining site are required for the antibody to bind a ligand as small as a trinucleotide. While these results, comparing the structures of the free Fab with the Fab complexed with the trinucleotide, are still preliminary, further refinement of the unliganded Fab structure will most likely not alter the overall conclusion that significant conformational change of the antibody CDRs takes place when the DNA is bound. This observation and the fact that BV04-01 binds oligonucleotides with a relatively low affinity lead the authors to conclude that the production of this antibody may have been stimulated by antigens other than DNA. In contrast, antibodies which are directed against proteinaceous antigens tend to bind with higher affinity, although evidence for an "induced fit" can also be found in the structures of the antibody-protein complexes.

The Anti-Lysozyme Complexes: Three separate structures of an Fab fragment in complex with hen egg white lysozyme have been determined: D1.3 (Amit et al., 1986), HyHEL-5 (Sheriff et al., 1987), and HyHEL-10 (Padlan et al. 1989). These structures have been recently reviewed (Davies et al. 1990). These three antibodies bind to different regions of lysozyme and have association constants in the nanomolar range: $1 \times 10^9\ M^{-1}$ (D1.3), $5 \times 10^9\ M^{-1}$ (HyHEL-10), and $2 \times 10^{10}\ M^{-1}$ (HyHEL-5) (Davies et al. 1990). The most salient feature of these complexes is the extraordinary complementarity between the two interacting surfaces. The surface complementarity is highlighted by the fact that there are almost no water molecules within the interface. Knobs on the surface of lysozyme are accommodated by holes within the antibody combining site. The total buried surface area of the interface is extensive, ranging from 690 $Å^2$ for D1.3, 721 $Å^2$ for HyHEL-10, to 746 $Å^2$ for HyHEL-5.

While there seems to be a slight preference for interaction with the H chain over the L chain, no particular CDR seems to be of paramount importance. All six CDRs contribute residues that make contact with the lysozyme in all three complexes. In addition, there is evidence for an induced fit of the two molecules rather than a strict "lock and key" mechanism. By comparing the conformation of the lysozyme in the HyHEL-5 complex with its uncomplexed structure, the peptide backbone in the region of proline 70 has moved by 1.7

Å (Sheriff et al., 1987). Unfortunately, the crystal structures of the uncomplexed Fabs have not been determined. However, crystal structures of the Fv domain of antibody D1.3 both free and in complex with lysozyme (Bhat et al., 1990) reveal that there is a small adjustment in the relative orientation of the V_H and V_L domains in the complex, similar to that observed for BV04-01. These studies reemphasize the point that flexibility on the part of both the antibody and the antigen may play a significant role in complex formation.

For antibodies that bind single stranded DNA (ssDNA) the potential for conformational change may be significant. ssDNA is flexible and does not possess a well defined secondary structure. This flexibility on the part of the DNA could be mirrored by flexibility of the hypervariable loops of the antibody. Without accurate knowledge of the structure the ssDNA adopts when bound to the antibody it is extremely difficult to predict with any degree of certainty which residues of the antibody will be interacting with the ssDNA. For antibodies that bind double stranded DNA (dsDNA) the potential conformational flexibility on the part of the antigen is reduced. One can generally assume that the dsDNA will be helical, however, the precise helical parameters can depend dramatically on the particular DNA sequence. Variations in the local helix geometry of DNA has been shown to occur in several crystal structures of short DNA oligonucleotides (Dickerson, 1987). In a model structure for poly(dG)•poly(dC) derived from the crystal structure of d(GGGGCCCC) (McCall et al. 1985) the helix assumes a conformation within the 'A' helix family. This may explain why poly(dG)·poly(dC) is more immunogenic than native 'B' DNA. The antibodies elicited by poly(dG)·poly(dC) can discriminate between this sequence and several related DNA polymers (Lafer and Stollar, 1984). For Jel 72 no binding is detected to poly(dG)·poly(dm^{5}C), poly(dGC)·poly(dGC), poly(dI)·poly(dC) or to poly(rG)·poly(dC) (Lee et al., 1984), suggesting that some base specific contacts are also being made.

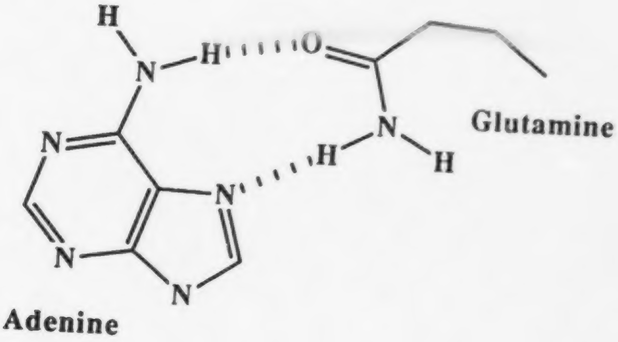
The mechanisms by which proteins are able to discriminate between different DNA bases has been partially elucidated by crystallographic analyses of protein-DNA complexes. These studies have been mostly conducted on repressor and activator proteins that bind specific DNA sequences. These proteins bind tightly to DNA as dimers and make several base specific contacts with the DNA. While the architecture of their DNA recognition domains can involve elements of protein secondary structure that are unlikely to occur in antibody combining sites, the specific interactions that are observed with the DNA bases may be applicable to antibody-DNA recognition.

I.4 DNA Binding Proteins

The Helix-turn-Helix Motif: Crystallographic studies of several prokaryotic and eukaryotic site specific DNA binding proteins have revealed that there are basic structural motifs that are particularly effective for the recognition of dsDNA. The helix-turn-helix (HTH) motif is a common structural fold employed by prokaryotic repressor and activator proteins. This motif was first recognized in the structures of the **cro** repressor protein from bacteriophage lambda (Anderson et al., 1981) and the **crp**, formerly CAP, activator protein from Escherichia coli (McKay and Steitz, 1981). Proteins that utilize this motif have an α-helix that projects from the main body of the protein. This helix fits into the major groove of DNA where it can make specific contacts with the edges of the DNA bases. More recently, structures of the bacteriophage lambda repressor bound to its operator sequence (Jordan and Pabo, 1988) and the related bacteriophage 434 repressor (Aggarwal et al., 1988) and **cro** (Mondragon and Harrison, 1991) proteins complexed with DNA have clarified the details ofthis interaction.

In the lambda and 434 repressor complexes similar contacts are made in the major groove between an adenine base and the side chains of either an asparagine or a glutamine residue (illustrated below). The assignment of particular amino acids to the recognition of specific DNA bases is complicated, however, by side chain-side chain interactions.

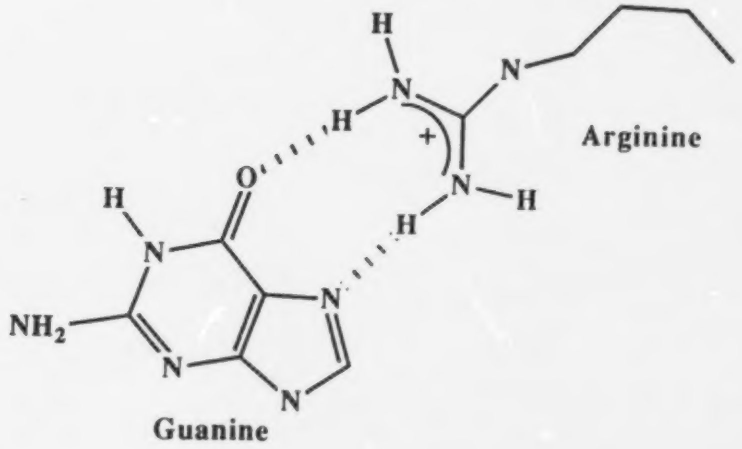
In both the lambda and 434 complexes the side chains of two glutamines form a network of hydrogen bonds linking an adenine base in the major groove with a phosphate of the deoxyribose backbone. As with the antibody-antigen complexes, there is surface complementarity between the two interacting molecules. There are base specific van der Waals contacts with the methyl group of thymine.



Both the lambda and 434 repressors interact extensively with the sugar phosphate backbone of DNA. These interactions may serve to position the recognition helix precisely within the major groove of the operator sequence. Many of these phosphate contacts are mediated by amino acids with small polar side chains or by direct interaction with the polypeptide backbone. Interestingly, lysines and arginines do not play a prominent role in these phosphate contacts. Predicting protein-nucleic interactions can also be complicated by the conformational flexibility of DNA. In the lambda repressor complex the DNA is in the B conformation. There is an average of 10.5 base pairs per turn, the DNA is relatively straight and the base pairs are nearly coplanar. The conformation of the DNA in the 434 repressor complex is quite different. The DNA is bent, the central base pairs are not coplanar, and there is a significant compression of the minor groove. Compression of the minor groove is generally unfavourable because it brings negatively charged phosphate groups close together. In the 434 repressor complex this compression is stabilized by the presence within the minor groove of the positively charged side chains of an arginine and a lysine.

The HTH motif is not restricted to prokaryotic proteins. The structure of the DNA binding homeodomain of the **engrailed** protein from Drosophila bound to DNA (Kissinger, et al., 1990) reveals that this eukaryotic protein also employs the HTH motif for recognizing DNA. The DNA in this complex is also a straight piece of B DNA with nearly coplanar base pairs. Again a recognition helix binds in the major groove and widens it. The resulting compression of the minor groove may be stabilized by the presence of the side chains of two arginine residues. Unlike the basic amino acids in the 434 repressor complex, these arginines also make base specific hydrogen bonds with the O2 of thymine bases in the minor groove.

EcoR1 Endonuclease: The structure of the restriction endonuclease EcoR1 complexed with DNA (McClarin et al., 1986; Kim et al., 1990) illustrates a different motif that utilizes α-helices for DNA recogniton. These helices are part of a 5-stranded parallel β-sheet structure termed the dinucleotide, or Rossman, fold. Unlike the helices of the HTH motif, the N-termini of these helices penetrate end on into the major groove. Binding of EcoR1 to DNA results in an unwinding of the DNA helix and a subsequent widening of the major groove allowing the N-termini of the α-helices to fit into the major groove. Base specific contacts are made by arginine and glutamic acid residues to guanine and adenine bases of the DNA. The side chain of one arginine forms a hydrogen bond with the N7 and O6 positions of a guanine (shown below), but a second arginine contacts two adjacent adenines at their N7 positions and the glutamic acid crosslinks these same two adenines by hydrogen bonding with their N6 atoms.



Zinc Fingers: The crystal structure of an eukaryotic DNA binding protein (Pavletich and Pabo, 1991) demonstrates how DNA can be recognized by a third major structural motif. The DNA binding domain from the mouse protein Zif268 belongs to the family of proteins homologous to transcription factor IIIa (TFIIIa) from **Xenopus** (Miller et al., 1985), containing a conserved sequence of two cysteines and two histidines that coordinate a zinc ion.

The zinc ion stabilizes the tertiary structure of a two-stranded antiparallel β-pleated sheet and an α-helix. Again it is the α-helix that binds in the major groove and makes base specific contacts with the DNA. The binding site for this particular zinc finger protein is guanine rich. Two types of interactions are observed between arginine side chains and guanine bases within the major groove. In both cases the guanido group of the arginine forms hydrogen bonds with the N7 and O6 atoms of the base. In three instances the arginine is also involved in a salt link with an aspartic acid, while in two cases the arginine side chain is positively charged.

A similar interaction between an arginine and a guanine is observed in the crystal structure of the DNA binding region of the rat glucocorticoid receptor complexed with DNA (Luisi et al., 1991). Unlike Zif268, the glucocorticoid receptor dimerizes upon binding and interacts with two succesive major grooves of the DNA. The zinc fingers of the glucocorticoid receptor differ from those belonging to the TFIIIa family in that the zinc ion is coordinated by four cysteines, rather than two cysteines and two histidines, but they recognize DNA in a similar manner. Once again base specific contacts are mediated by an α-helix within the major groove. Other than the arginine-guanine interaction there is also a van der Waals contact between a thymine methyl group and a valine residue.

These structures emphasize the central role that an α-helix plays in many protein-nucleic acid interactions. The occurrence of recognition helices in otherwise disparate structural motifs underlines the inherent complementarity between the cylindrical shape of the α-helix and the helical saddle-shape of the major groove of B DNA. Early model building studies (Seeman et al., 1976) had recognized this shape complementarity, but had also suggested that a two-stranded antiparallel β-ribbon should also be able to fit into the major groove and contact the exposed edges of the base pairs.

Proteins that use β-strands to bind DNA: There are two families of proteins which interact with DNA via β-strands. Members of the HU family, which includes HU, TF1 and the integration host factor (IHF) from E. coli, interact with DNA through the minor groove. In the crystal structure of the HU protein from B. stearothermophilus, a two-stranded β-ribbon is proposed to insert into and expand the minor groove, resulting in an appreciable bending of the DNA (White et al., 1989). According to the proposed model, conserved lysine and arginine residues make contact with the bases and phosphodiester backbone of the DNA.

Members of the second class of proteins that use β-strands to bind DNA are typified by the MetJ repressor (Rafferty et al., 1989). These proteins recognize DNA by insertion of a two-stranded antiparallel β-ribbon in the major groove. The β-ribbon is part of a ribbon-helix-helix motif that serves to position the β-ribbon within the major groove (Phillipps, 1991). Sequence specific hydrogen bonds are made by a lysine residue to a guanine base, and by a threonine to an adenine base. Another member of this class of proteins is the Arc repressor from the Salmonella bacteriophage P22 (Knight et al., 1989). The solution structure of the Arc repressor has been determined by NMR spectroscopy and a model for its interaction with DNA proposed (Breg et al., 1990). The Arc repressor makes specific contacts with the DNA by binding of a two-stranded N-terminal anti-parallel β-sheet in the major groove of the DNA.

I.5 Conclusions

Taken together these structures illustrate several diverse structural motifs that proteins employ to bind specifically to DNA. While conserved contacts between amino acid side chains and DNA bases are observed in several different complexes, it is apparent that the specific interactions observed are dictated by the precise nature of the particular protein DNA interface. Base specific contacts can be influenced by side chain-side chain interactions, by side chains that contact more than one base, by hydrogen bonding through bound water molecules, or by interactions with the polypeptide backbone.

The application of the results derived from these studies to DNA recognition by antibodies is further complicated by the nature of the antibody combining site, which is almost entirely comprised of β-strands, turns and loops. This architecture places constraints on the structural motifs that antibodies can employ to recognize DNA. Since antibodies lack any significant helical structure, any model for DNA binding based on an α-helix will not be relevant. The need for a divalent metal ion in DNA binding has not been observed with antibodies, diminishing the possibility that metal nucleated substructures play a role in recognition. We thus had reason to believe that the structural motifs antibodies employ in DNA recognition would differ to some degree from the previously observed motifs and undertook the determination of the three-dimensional structures of several different DNA binding antibody Fab fragments using the techniques of X-ray crystallography.

A complete treatment on the theory and practice of X-ray crystallography is outside the purview of this thesis. Readers interested in an introduction to these techniques and their application to the determination of protein structures are directed to the text by T.L. Blundell and L.N. Johnson (1976). X-ray crystallography is the only method presently known that can reliably determine high resolution three dimensional structures of molecules as large as antibody Fab fragments. A major disadvantage, however, is that large diffraction quality single crystals of the molecule of interest must first be obtained. Antibodies belonging to the IgG class can be routinely incubated with proteolytic enzymes to produce free Fab fragments. Fab fragments are generally easier to crystallize than entire antibody molecules, and are thus more amenable to investigation by diffraction methods. Our experience has shown that digestion of IgG antibodies with papain produces a heterogenous mixture of Fab fragments. In order to grow suitable crystals these Fab fragments must first be purified to isoelectric homogeneity. Chapter II describes the purification protocols that we have employed to obtain the required homogeneity and the effects impurities have on the crystallization process.

In Chapter III the structure determination and refinement of two separate crystal forms of Jel 72 and one form of Jel 318 is described. These structures were determined by the technique of molecular replacement (Rossman, 1990) which takes advantage of the fact that most of the structure of an Fab fragment is conserved and is expected to be very similar to previously determined Fab fragment structures.

The observation that some of the hypervariable loops of antibody combining sites belong to discrete structural classes (Chothia and Lesk, 1987) and can therefore be modelled based on known structures, is addressed in Chapter IV. We constructed models for three DNA binding Fvs and compared these models with the structures of the combining sites as determined by X-ray crystallography. The modelled structures proved to be reasonably accurate and the modelling protocol was then applied to several DNA binding Fv domains whose amino acid sequences had been determined but for which no crystallographic structures are available.

An examination of the shape and electrostatic surface potentials of both the modelled and crystallographically determined structures has enabled us to propose testable hypotheses for DNA recognition by these antibodies. These DNA binding models are described in Chapter V.

The fact that we have not obtained any crystallographic data on the Fab-nucleic acid complexes must be reiterated. Lacking these data, the models that we propose represent speculation on our part regarding possible interactions between the antibodies and their nucleic acid antigens. As the co-crystal structure of the BV04-01 trinucleotide complex illustrates, the antibody combining site may undergo extensive conformational change upon binding nucleic acids. This potential for an induced fit of the combining site to the antigen severely restricts our ability to propose specific interactions. Nevertheless, certain features of the combining sites, namely the disposition of the hypervariable loops with respect to one another and the location of regions of positive electrostatic surface potential, may be important factors in the recogniton of nucleic acids by these antibodies. The models that we propose are, thus, intended to assist in the design and execution of future experiments to further advance our understanding of antibody-DNA recognition.

II. Purification and Crystallization

II.1 Introduction

Structural studies on antibody Fab fragments were initiated in our laboratory with the mouse autoimmune antibody Hed10, which is highly specific for single stranded poly(dT). We began this work in collaboration with the laboratory of Dr. Jeremy S. Lee. At the beginning of our collaboration, the production of the Fab fragments from Hed10 was performed in Dr. Lee's laboratory and the Fab preparations sent to us for use in the crystallization trials. When a batch of Fabs was received that failed to produce crystals under conditions that gave crystals from the preceding batch, we began investigating the purity of the Fab preparations using analytical isoelectric focussing (IEF). The Fab preparation that produced large single crystals was homogenous, containing a single isoelectric species, but in the succeeding preparations several different isoelectric bands were also detected on the IEF gel. We then arranged for Dr. Lee to send us the intact antibodies so we could produce the Fab fragments ourselves.

The microheterogeneity of the Fab preparations could be minimized by adjusting the conditions of the papain digestion. We found that longer incubation times (4 - 5 hours) and higher papain concentrations (0.04 - 0.08 mg/ml) would minimize the amount of contaminating Fab species with higher isoelectric points that were produced. Another important factor in the reproducibility of the digestions was the freshness of the papain. Older papain solutions tended to produce unpredictable distributions of Fab species, often producing the Fab species of interest in lesser amounts than the species that would not crystallize. Even under optimum digestion conditions, however, two major isoelectric forms of Fabs were invariably produced. The difference in pI of these species was on the order of 0.2 pH units, which presented us with a challenging purification problem. Although we achieved some success purifying these species using a preparative IEF protocol, the technique was cumbersome and expensive and did not produce the yields of pure Fab fragments that we required for the crystallization experiments. An alternate purification scheme was then devised utilizing conventional column chromatography on a DEAE cellulose column. This purification protocol greatly facilitated the preparation of concentrated solutions of isoelectrically pure Fab species. Large, single crystals could then be reproducibly grown and the structure determination of Hed10 was able to proceed apace (Cygler, et al., 1987).

This chapter describes similar purifications of the Fab preparations of Jel 72 and Jel 318, and the effects that the presence of different isoelectric species had on the crystallization experiments. The space group determination and initial crystallographic characterization of two crystal forms of Jel 72 and one form of Jel 318 are also described. A version of this chapter has been published (Boodhoo, et al, 1988).

II.2 Purification of Jel 72

Crystallization experiments with the unpurified mixture of Jel 72 Fab fragments were initially performed in our laboratory by Amechand Boodhoo, and microcrystals unsuitable for diffraction analysis had been obtained. As a result of our experiences with Hed10, we examined the heterogeneity of the Jel 72 Fab fragments by analytical IEF. Several trial digestions with papain (Sigma) were performed in which the papain to IgG weight ratios were varied between 1:25 and 1:400. The ratio that resulted in the least heterogeneity was chosen for preparative digestions. A papain to IgG weight ratio of 1:50 was routinely used with fresh papain preparations. Older enzyme solutions required higher concentrations of papain (1:25 weight ratio) to produce similar distributions of Fab isoelectric species. Incubations were typically for 4 hours at 37°C in 0.1-0.25 M NaCl, 20 mM Tris-HCl, pH 7.4, 2 mM EDTA, and 2-25 mM β-mercaptoethanol.

Papain is a thiol protease. It contains an active site cysteine residue that forms a covalent thioester bond between an acyl group on the substrate and the sulphydryl group of the cysteine. The β-mercaptoethanol is required in the digestions to reduce the active site sulphydryl group allowing the enzyme to continue on to another round of proteolysis. The papain was inactivated with iodoacetamide, which blocks the sulphydryl group of cysteine residues, and the digests loaded onto isoelectric focussing gels or dialyzed against appropriate column loading buffers for purification.

Cleavage of Jel 72 with papain results in two major Fab species, pI 8.6 and pI 8.8, and several minor ones with pIs > 9.0 (Figure II-1). The higher pI species predominate if very short digestion times or low concentrations of papain are employed. These higher pI bands are exclusively produced if pepsin (Sigma) is used in the digestion. At higher papain concentrations the pI 8.6 and pI 8.8 species predominate. Best results were obtained at relatively high concentrations of papain to IgG (1:25 to 1:50 weight ratio).

The purification of these Fab fragments was initially achieved by preparative IEF as described in the Pharmacia product manual (Phamacia, Inc.) and subsequently by anion exchange on a QA52 column (Whatman). In order for the Fab fragments to bind, they must be loaded onto the column at a pH above their isoelectric points. The pIs of the Jel 72 Fabs are too high to bind to DEAE cellulose, which loses its charge above pH 8.0 due to titration of the positively charged tertiary amino groups on the column. If the digests were buffered to pH 10.0 with 20 mM 3-cyclohexylamino-1-propane sulfonic acid (CAPS) as the buffer, the Fab fragments would bind weakly at this pH to the QA52 column and could be eluted with a 0.0 - 0.1 M NaCl gradient (Figure II-2). The pI 8.8 Fab species was the major isoelectric species produced in the digestion and eluted from the column at an NaCl concentration of approximately 40 mM. This species was primarily used in the crystallization experiments. The Fab fragments with isoelectric points above 9.0 eluted in the flow through or appeared as a small peak preceding the pI 8.8 species. The pI 8.6 species eluted from the column immediately after the pI 8.8 species.

As the Fabs were eluted from the column they were immediately neutralized with Tris-HCl, pH 7.4, and the β-mercaptoethanol concentration raised to 2 mM. Column fractions were analyzed on analytical IEF gels and those fractions that contained only the pI 8.8 species were pooled and concentrated and used in the crystallization trials. The purified pI 8.6 Fab species did not produce crystals under the same conditions that yielded crystals with the purified pI 8.8 Fab species.

The Fc fragments could be eluted from the column with buffer containing 0.5 M NaCl. The Fc fragments of Jel 72 seemed to be sensitive to further proteolysis by papain and did not always focus as discrete bands on the IEF gels (see Figure II-1, compare Jel 72 with Jel 318 Fc fragments).

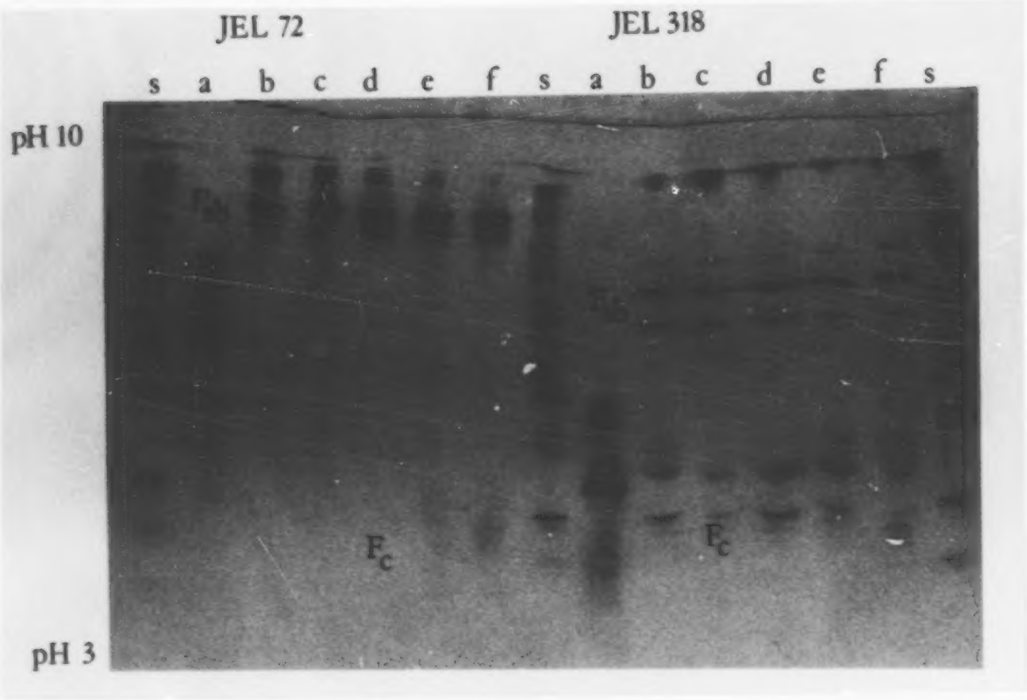


Figure II-1: Analytical IEF Gel of Digests of Jel 72 (left) and Jel 318 (right) immunoglobulins. Descriptions were performed as described in the text with the following papain:IgG weight ratios: 1:400 (lane b), 1;200 (lane c), 1:100 (lane d), 1:50 (lane e), and 1:25 (lane f). Lane a contains undigested IgG. Lane S, IEF standards (Pharmacia): β-lactoglobulin A, pI 5.2; bovine carbonic anhydrase B, pI 5.85; human carbonic anhydrase B, pI 6.55; horse myoglobin, pI 6.85 and 7.35; lentil lectins, pI 8.15, 8.45, and 8.65; and trypsinogen, pI 9.3. The positions of the Fab fragments in lanes b-f are indicated by labels to the right of the bands, Fc fragments are labelled at their approximate positions on the gel.

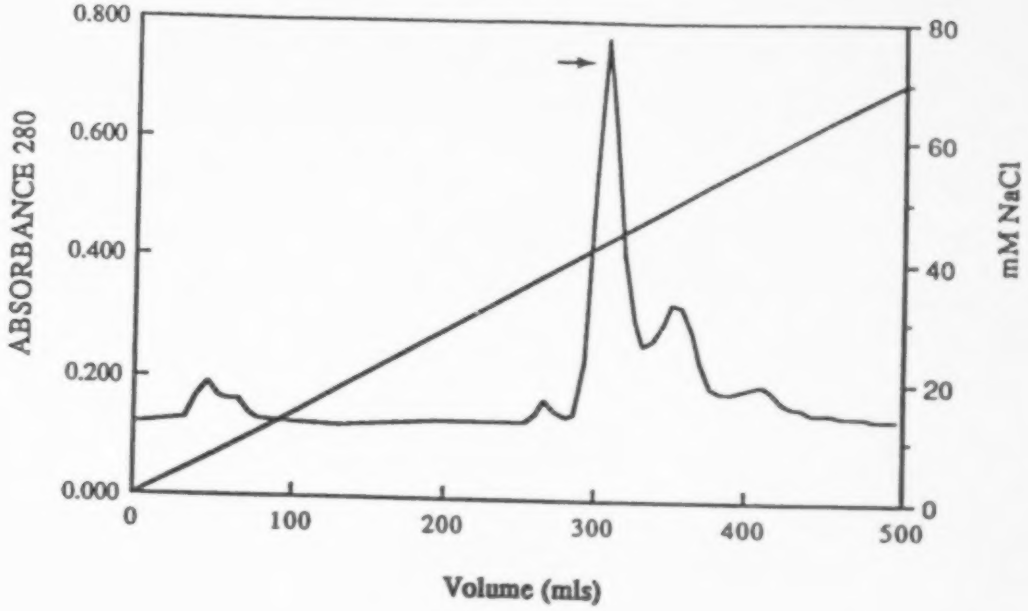


Figure II-2: Purification of Jel 72 Fab Fragments. Elution profile from a Qa52-cellulose anion exchange column showing separation of Fab peaks. Jel 72 IgG papain digests were dialyzed vs pH 10.0 CAPS buffer prior to loading onto the column and eluted with a 0.0 - 0.1 M NaCl gradient. The first large peak (arrow) corresponds to the pI 8.8 Fab species used in the crystallization experiments..

II.3 Crystallization of Jel 72 in Crystal Form (I)

The Fab fragments of Jel 72 were crystallized by the technique of hanging drop vapour diffusion (McPherson, 1976) at room temperature. The purified pI 8.8 species of Jel 72 Fab, 9 mg/ml in 28% saturated ammonium sulphate (AS) and 25 mM sodium acetate, pH 4.2, was equilibrated through the vapour phase against 56% AS, 50 mM sodium acetate, pH 4.2. Microcrystals obtained from the mixture of Jel 72 Fabs and a crystal obtained from the purified pI 8.8 species are shown in Figure II-3. The crystals can be grown to as large as 1.0 x 0.6 x 0.4 mm and diffract to about 2.6 Å.

The crystallographic space group and unit cell dimensions were determined from 12° precession photographs. The photographs displayed **mm** symmetry about each of the three mutually perpendicular principal axes and showed extinctions along h00, h=2n + 1, 0k0, k=2n + 1 and 00l, l=2n + 1, indicating that the Fab fragment crystallizes in the orthorhombic space group $P2_12_12_1$, with refined unit cell parameters a=94.63, b=102.60, c=92.42 Å. Assuming that there are two Fabs of M_r 50,000 in the asymmetric unit yields a V_m of 2.40 $Å^3$/dalton, which is in the range observed for other protein crystals (Matthews, 1968).

These results dramatically illustrate the need for isoelectric homogeneity in order to obtain usable crystals of Jel 72 Fab fragments. Even minor amounts of contamination by the pI 8.6 species resulted in crystals that were badly intergrown and unsuitable for data collection.

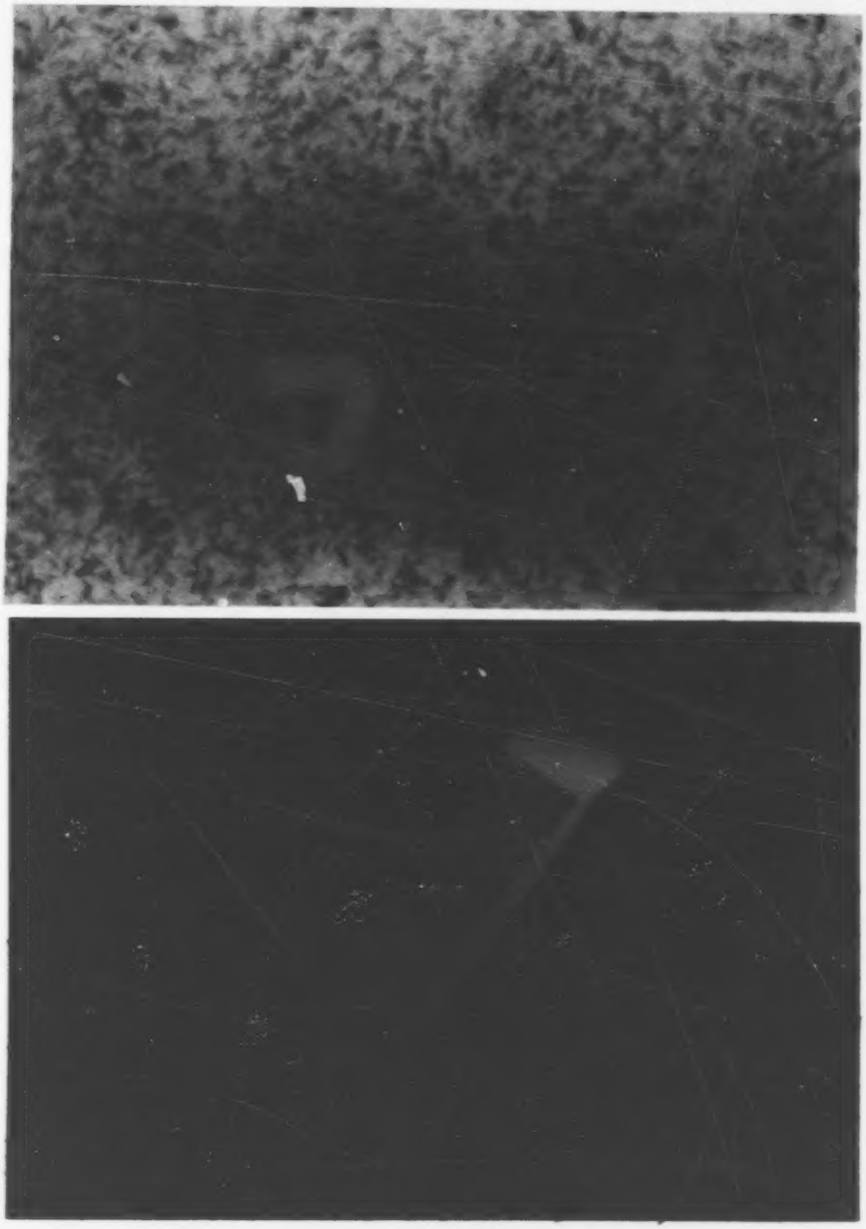


Figure II-3: Crystals of Jel 72 Fab Fragments. Microcrystals obtained from unpurified mixture of Fab isoelectric species (top), and a large single crystal grown from purified pI 8.8 Fab species (bottom). The crystal shown has approximate dimensions of 0.7 x 0.3 x 0.2 mm. The magnification used in both photographs is identical

II.4 Crystallization of Jel 72 in Crystal Form (II)

After we had obtained the crystals of Jel 72 with two Fabs in the asymmetric unit we redoubled our efforts to grow crystals of Jel 72 complexed with various double stranded oligonucleotides. Co-crystallization experiments were performed with a number of double stranded oligonucleotides including:

i) $d(G_4)\cdot d(C_4)$
ii) $d(G_5)\cdot d(C_5)$
iii) $d(CG_4C)\cdot d(GC_4G)$
iv) $d(CG_6C)\cdot d(GC_6G)$
v) $d(CG_8C)\cdot d(GC_8G)$
vi) $d(AAG_6AA)\cdot d(TTC_6TT)$
vii) $d(ATG_8T)\cdot d(TAC_8A)$
viii) $d(CG_9CA)\cdot d(GC_9GT)$
ix) $d(CG_{10}C)\cdot d(GC_{10}G)$

Although small, football shaped, highly birefrigent crystals were reproducibly obtained from solutions containing the pI 8.8 Fab species of Jel 72 and the eleven base pair oligonucleotide $d(ATG_8T)\cdot d(TAC_8A)$, these crystals did not diffract and could not be made to grow larger than approximately 0.1 x 0.05 x 0.05 mm.

It was in these co-crystallization experiments that a crystal was grown to dimensions of 0.15 x 0.15 x 0.1 mm from a solution containing 3.5 mg/ml of the pI 8.6 Fab species, 0.3 mM of the oligonucleotide $d(CG_9CA)\cdot d(GC_9GT)$, 100 mM Piperazine-

N,N'-bis[2-ethanesulfonic acid] (PIPES), pH 6.1, 5 mM $MgCl_2$, 200 mM NaCl, and 10% polyethylene glycol (PEG) 8000.

The crystal was mounted and the space group and unit cell dimensions determined by inspection of the reciprocal lattice using a multiwire area detector from San Diego Multiwire Systems (Xuong et al, 1978). These crystals also belong to the orthorhombic space group $P2_12_12_1$, with refined unit cell dimensions of a=90.56, b=140.71, c=37.44 Å, but contain only one Fab in the crystallographic asymmetric unit. The structure determination and refinement at 2.9 Å resolution of this crystal form is described in the next chapter. Inspection of the electron density of this refined structure and consideration of the crystal packing interactions indicates that the crystal does not contain any bound oligonucleotide. This structure is of interest, however, because the manner in which it packs in the crystal is similar to that of Jel 318.

II.5 Purification and Crystallization of Jel 318

Digestion of the Jel 318 with papain produced two major Fab species with isoelectric points of 7.6 and 7.8 and one minor species with a pI of 8.0 (Figure II-1). In contrast with the earlier results for Hed10 (Cygler et al., 1987) and Jel 72, the distribution of products of the digestion were relatively insensitive to papain concentration. The digestion was stopped with iodoacetamide and dialyzed versus 10 mM Tris-HCl, pH 8.4, prior to loading the digest onto a DEAE-Sephacel (Pharmacia) column with the same buffer. The Fabs were eluted from the column with a 0.0 - 0.25 M NaCl gradient (Figure II-4). Although this protocol did not achieve a clean separation of the two Fab fragment peaks, purified pI 7.8 Fabs could be obtained by pooling only those fractions that were at the front of the peak and performing the purification again as necessary.

The crystallization behaviour of the Jel 318 Fab fragments also differed from Hed10 and Jel 72. Both of the purified Fab isoelectric species could be crystallized by vapour diffusion at room temperature. The pI 7.8 species crystallized from 12% PEG 8000, 50 mM NaCl and 10 mM Tris-HCl, pH 7.8. Precession photographs to 12° showed the same symmetry and extinctions as those for Jel 72, indicating that the space group was again $P2_12_12_1$, but with refined unit cell dimensions a=82.89, b=139.190, c=40.96 Å. These crystals diffract poorly beyond 3.0 Å with usable data to about 2.8 Å resolution.

The pI 7.6 species yielded crystals with nearly identical unit cell dimensions under similar conditions. When a pure solution of the pI 7.8 species was mixed with small amounts of the pI 7.6 species the crystals obtained were inferior in terms of both their size and quality.

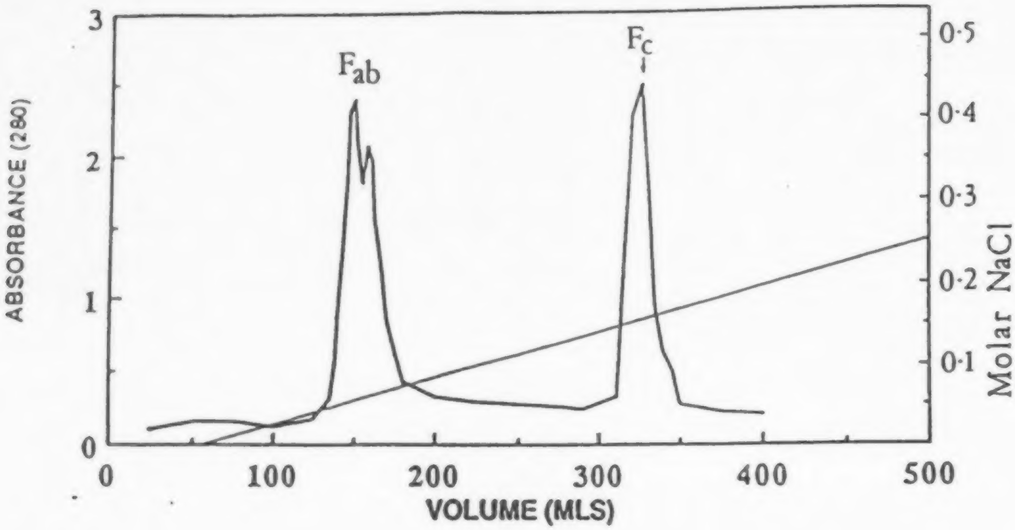
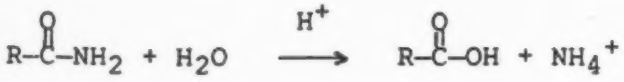


Figure II-4: Purification of Jel 318 Fab Fragments. Elution profile from a DEAE-cellulose anion exchange column of Jel 318 papain digests. The digests were loaded onto the column at pH 8.4 and eluted with a 0.0 - 0.25 M NaCl gradient. The first two peaks correspond to the Fab fragments while the third peak is the Fc fragment.

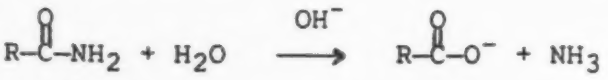
II.6 Discussion

The results presented in this chapter clearly indicate that crystals suitable for X-ray diffraction studies can be obtained from preparations that contain a single isoelectric species of Fab fragments. Both the nucleation and growth of crystals can be profoundly affected by the presence of relatively minor amounts of contaminating isoelectric species. In the case of Hed10, crystals could not be grown from the mixture of Fab species (Cygler et al., 1987). For Jel 72 only microcrystals could be obtained and, although crystals could be grown from the mixture of Jel 318 Fabs, they grew slowly and to a smaller size than the crystals obtained from the purified Fab preparations. A fourth antibody studied in our laboratory, Jel 150, which is specific for the left-handed conformation of poly(dGC), also produced better crystals upon further purification of the Fabs. Two major Fab species with very high (pI > 9.0) and similar isoelectric points were produced from the papain digestion of Jel 150 IgG. Crystals grown from this mixture are extremely thin plates, whereas crystals grown from the purified isoelectric species are considerably thicker. Taken together, these results suggest that while isoelectric purity is not the only factor governing success or failure in a crystal growth experiment, it can be a very important one. In all four cases studied, with the pure isoelectric species crystal nucleation occured readily and the crystals grew to a relatively large size in a period of a few weeks.

Papain digestion of an intact IgG usually produces at least two major Fab species. There is no way of knowing, **a priori**, which, if any, of these species will crystallize. Crystallization trials should be carried out on each of the purified species to obtain the best quality crystals. Freshly purified Fabs invariably gave the best results in the crystallization experiments. In the case of Jel 72, a pure solution of the pI 8.8 species obtained by preparative IEF and which initially gave large single crystals, failed to crystallize after a period of storage at 4°C of about 2 weeks. At this time significant amounts of the contaminating pI 8.6 band could be detected by analytical IEF. This conversion may be due to protease contamination or due to the deamidation of glutamine or asparagine residues. Amides can react with water in the presence of either strong acid or strong base to produce the corresponding carboxylic acids:



$$R\text{-}\overset{O}{\overset{\|}{C}}\text{-}NH_2 + H_2O \xrightarrow{H^+} R\text{-}\overset{O}{\overset{\|}{C}}\text{-}OH + NH_4^+$$



$$R\text{-}\overset{O}{\overset{\|}{C}}\text{-}NH_2 + H_2O \xrightarrow{OH^-} R\text{-}\overset{O}{\overset{\|}{C}}\text{-}O^- + NH_3$$

The deamidation of asparagine or glutamine residues may be a particular concern for Jel 72 Fab fragments, since their purification required basic conditions (pH 10.0) and the crystals grow under acidic conditions (pH 4.2).

Analytical IEF is an invaluable technique for analyzing the microheterogeneity produced by digestion of an IgG with papain or for monitoring sample purity during storage. Mixtures of immunoglobulin Fab fragments with isoelectric points in the pH range 7 to 9 can be easily purified on an appropriate anion exchange column. The possibility of obtaining large single crystals suitable for X-ray diffraction studies is likely to be greater with the use of isoelectrically homogenous Fab species in the crystallization experiments.

III. Structure Determination and Refinement

III.1 Introduction

Obtaining large, single crystals that diffract to high resolution is only the first, and often most formidable, step in the determination of the structure of a biological macromolecule. Once the crystallographic space group has been determined and the data collected, the next hurdle is what crystallographers term the phase problem. The nature of the phase problem can be appreciated by inspection of the electron density equation:

$$\rho(xyz) = 1/V \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl)\exp i\alpha(hkl)\exp{-2\pi i(hx+ky+lz)}$$

The data that are measured represent only the amplitudes, F(hkl) of the diffracted X-rays, while the calculation of electron density, $\rho(xyz)$ requires that both the amplitudes and the phases, $\alpha(hkl)$ be known. All information regarding the phases of the diffracted beams is lost when the data are measured. In order for a crystal structure determination to be successful this missing phase information must be reconstructed.

One technique for obtaining the missing phase information that is commonly employed for the determination of the structures of biological macromolecules is the method of isomorphous replacement. This technique was first applied successfully to the structure determination of hemoglobin

(Green et al., 1954) and many macromolecular structures have since been successfully determined using this method. Isomorphous replacement involves introducing a heavy atom into a crystal in such a way that the only difference in the electron density is a peak at the site where the heavy atom binds. For the technique to be a success the unit cell dimensions of the heavy atom derivative crystal must not differ significantly from the dimensions of the native crystal, and the two crystals must have identical structures for their protein components. That is, the native and derivative crystals must be isomorphous.

The most common method for obtaining isomorphous derivatives of crystals of biological macromolecules is by soaking the native crystals in solutions containing the heavy atom compounds. Crystals of biological macromolecules often contain large channels of mother liquor through which the heavy atom can diffuse and then bind to groups on the surface of the macromolecule in the crystal.

A useful heavy atom will contain many more electrons than the atoms generally found in biological macromolecules (C,N,O,S). This fact allows the positions of the heavy atoms within the crystal to be determined by the use of the Patterson function (Patterson, 1934), $\mathbf{P}(uvw)$:

$$\mathbf{P}(xyz) = 2/V \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=0}^{\infty} F(hkl)^2 \cos 2\pi i(hu+kv+lw)$$

Since the Patterson function is calculated using only the squares of the amplitudes of the diffracted X-rays, it can always be computed from a set of intensity data. A Patterson map represents all the interatomic vectors of the molecules in the crystal. The Patterson function is, thus, centrosymmetric, since for every interatomic vector between atom **i** and atom **j** there is a corresponding vector in the opposite direction between atom **j** and atom **i**. In general, the size of a peak in a Patterson map that represents a vector from atom **i** to atom **j**, has a weight that is proportional to $\mathbf{Z}_i\mathbf{Z}_j$, where atom **i** contains $\mathbf{Z}_i$ electrons and atom **j** contains $\mathbf{Z}_j$ electrons. Thus, a difference Patterson function calculated by subtracting the native Patterson from a heavy atom derivative Patterson will contain large peaks that correspond to the interatomic vectors between the bound heavy atoms.

The positions of the heavy atoms can be determined by inspection of the Harker sections of the Patterson map. The Harker sections contain the interatomic vectors of the atoms that are related by the space group symmetry elements, and the peaks found in these sections can be used to infer the locations of the heavy atoms. Once the positions of the heavy atoms have been determined they are used to calculate a set of phases that can be used to estimate the phases for the rest of the structure.

There will be an ambiguity in these phases, however, if the information from only a single heavy atom derivative is available. This phase ambiguity is best illustrated through the Harker construction (Harker, 1956) for the phase calculation, which is illustrated in Figure III-1. If $\mathbf{F}_{PH}$ is the structure factor of the heavy atom, $\mathbf{F}_P$ the structure factor for the native protein, and $\mathbf{F}_H$ is the contribution of the heavy atoms to the structure factor of the derivative, then the situation depicted graphically in Figure III-1 illustrates that there are two possible solutions for $\mathbf{F}_P$ which are symmetrically disposed about $\mathbf{F}_H$. Only when the vectors $\mathbf{F}_P$ and $\mathbf{F}_H$ are colinear is there no ambiguity in the phase. Additional phase information is required from other independent derivatives to resolve the ambiguity that exists in the estimate of the phases and to indicate which of the two possible solutions for $\mathbf{F}_P$ is the correct one. The technique of multiple isomorphous replacement uses the phases provided by two or more derivatives to calculate an initial electron density map from which the structure of the macromolecule can be interpreted. Readers interested in a more detailed description of the determination of protein crystal structures by the method of multiple isomorphous replacement, and the treatment of errors in the method, are directed to the text by Blundell and Johnson (1976).

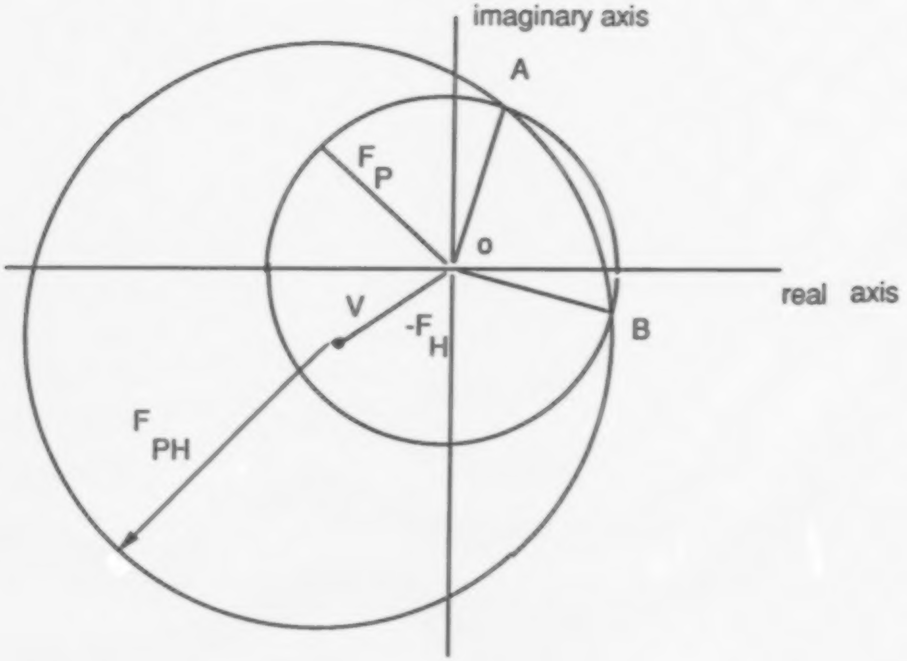

imaginary axis
A
F P
o
real axis
V
-F H
B
F PH

Figure III-1: The Harker construction for phase calculation using a single isomorphous derivative. Two circles of radii $\mathbf{F}_P$ and $\mathbf{F}_{PH}$ are drawn with their centers on O and V, respectively. The vector **V** represents $-\mathbf{F}_H$. The vectors OA and OB represent the two possibilities for $\mathbf{F}_P$.

There is another method that can provide the missing phase information if the molecule of interest is homologous to a molecule of known structure. The technique of molecular replacement (Rossman, 1990) also takes advantage of the information contained within the Patterson function. All Patterson maps will contain peaks near the origin that represent the interatomic vectors between the atoms of an individual molecule. One set of these intramolecular vectors will be present for each orientation of the molecule in the crystallographic cell. Rotation function calculations compare the Patterson maps of the molecule in the crystal and an imaginary crystal containing the homologous molecule of known structure. A peak in the rotation function corresponds to the rotations required to place the observed Patterson map in an orientation that maximizes the agreement near the origin with the model Patterson map. This will be the orientation required to orient the known model like the unknown structure. This oriented model is then positioned within the crystal by means of a translation search.

Two types of translation search have been employed in our laboratory. The first utilizes a brute force approach that positions the oriented models at successive points within the crystallographic cell. It then calculates a correlation coefficient between the observed structure factor amplitudes and those derived from the oriented and translated model (Fujinaga and Read, 1987). When the oriented model is at the correct position the correlation coefficient should reach a maximum value.

The second approach to determining the translation vector for the oriented model takes advantage of prior phasing information obtained from isomorphous replacement. The phased translation function (Read and Schierbeek, 1988) computes the reciprocal space equivalent of the correlation between the electron density calculated with the phases obtained from isomorphous replacement and the electron density calculated from the oriented and translated model. A maximum overlap of these two electron density maps gives the correct translation vector.

When the model structure is correctly positioned within the crystallographic unit cell it provides phases that are used to calculate an initial electron density map. Inspection of this electron density map can reveal areas where the molecule of interest differs from the molecule of known structure and the model adjusted accordingly.

The conserved structure of the β-sheet core of antibody Fab fragments facilitates the use of molecular replacement for determining their crystal structures. Unfortunately, the interpretation of the rotation and translation search results is complicated by the variation in the dispositions of the variable domain pair relative to the constant domain pair (the elbow angle), and of the heavy chain relative to the light chain (the pseudotwofold angle). In our experience we have found the results to be more interpretable if we perform the calculations on the variable and constant domain pairs separately and use a variety of known Fab fragments as models.

The use of this technique in our laboratory was initiated by Dr. Mirek Cygler, who determined the rotation and translation parameters for both Hed10 and Jel 318. For the determination of the orientations of the variable and constant domain pairs in the solution of the structure of Jel 72 with two Fab fragments in the crystallographic asymmetric unit, I am indebted to Dr. Alastair Muir, whose interpretation of the rotation function calculations proved to be largely correct.

The determination of the structure of Jel 72 in crystal form I was complicated by having two Fab fragments in the asymmetric unit. Rotation and translation searches were carried out with individual domain pairs that represented only 1/4 of the contents of the asymmetric unit. This proved to be just enough to discern the correct solution above the background noise. Although Jel 318 and the second crystal form of Jel 72 each contained only one Fab fragment in the asymmetric unit, the determination of their structures was made more difficult by an unusual variation in the positioning of the Fab fragment that resulted in crystals with similar unit cell dimensions having different crystal packing interactions.

Once the model Fab domains have been positioned correctly within the unit cell the structure must be refined and the amino acid sequence of the hypervariable loops fitted to the electron density of the antibody combining region. An initial rigid body refinement optimizes the positions of the V_L, V_H, C_L and C_H domains. In subsequent rounds of refinement the positions of the atoms are varied to maximize agreement with the observed electron density while maintaining reasonable stereochemistry. The method that I have employed to refine the Fab fragment crystal structures is that of simulated annealing (Kirkpatrick et al., 1983), which employs molecular dynamics to explore the conformational space of the molecule, as implemented in the X-PLOR system of programs (Brünger et al., 1989). Crystallographic refinement using X-PLOR is accomplished through molecular dynamics simulations that are performed by solving Newton's equations of motion (Verlet, 1967) with forces on the atoms that are derived from an empirical potential energy that describes stereochemical and nonbonding interactions (Karplus and McCammon, 1983). The molecular dynamics simulations are included in the crystallographic refinement by adding the effective potential energy, E_{SF} :

$$E_{SF} = S \sum_{hkl} [\,|F_{obs}(hkl)| - |F_{calc}(hkl)|\,]^2$$

to the empirical potential energy. E_{SF} describes the differences between the observed and calculated structure factor amplitudes, and is identical to the function employed in conventional least-squares refinement methods (Jack and Levitt, 1978).

For all three of the Fab fragment structures that have been determined in our laboratory, the interpretation of the electron density was hindered by the fact that the X-ray analyses preceded the determination of the length and amino acid sequences of the hypervariable regions of the antibodies. As the sequences were determined in Dr. Jeremy Lee's laboratory they were fitted to the density and the refinement was able to proceed. The poor quality of the electron density in the flexible solvent-exposed regions of the loops and the medium resolution of the data makes it unlikely that accurate descriptions of the antibody combining sites could have been derived without first knowing the amino acid sequences of the hypervariable loops.

When the sequences were determined the structures of both Jel 72 and Jel 318 were only partially refined. We then modelled the structures of their Fv regions using the techniques described by C. Chothia and A.M. Lesk (1987). After the refinements were completed we were in a position to compare the crystal structures with the model structures that we had derived. The success of this modelling protocol is addressed in the next chapter.

III.2 Structure Determination and Refinement of Jel 72 (formI)

The initial attempts at determining the structure of Jel 72 with two Fab fragments in the crystallographic asymmetric unit by molecular replacement were not successful. At this time a search for heavy atom derivatives was begun and three useful ones found. When we later reexamined the rotation function calculations and discovered peaks which we believed to be the correct solutions, the phasing information provided by these derivatives was extremely valuable for positioning the oriented domain pairs using the phased translation search (Read and Schierbeek, 1988). The use of this technique with Jel 72 demonstrates the efficacy of this approach but also illustrates some of the problems that can arise when the orientation angles are in error. The phased translation search conducted with a misoriented variable domain pair produced a result that was close to, but not precisely the same as, the correct translation. Fortunately, we were able to arrive at the correct solution during the structure refinement by performing rigid body refinement on the mispositioned domains while constraining the other domains to remain in their correct positions.

Data Collection:

As indicated in Chapter II, this crystal form of Jel 72 is obtained from 54% saturated ammonium sulphate and 50 mM sodium acetate pH 4.2. The crystals belong to space group $P2_12_12_1$, with a=94.63, b=102.60, and c=92.42 Å. Native crystals of Jel 72 were stabilized by equilibration against reservoir solutions of 65% saturated ammonium sulphate, 100 mM sodium acetate, pH 4.2, prior to data collection. Three heavy atom derivatives were obtained by soaking native crystals in ammonium sulphate stabilizing solutions containing 2 mM $UO_2(NO_3)_2$ for 1 week, 0.25 mM $K[Au(CN)_2]$ for 6 days, and in 0.02 mM $K_2[Pt(CNS)_6]$ for 4 days. Intensity data from native and derivative crystals were collected with a multiwire area detector from San Diego Multiwire Systems, using CuKα radiation. Native data were collected to 2.7 Å, while derivative data were collected to lower resolutions. The data were corrected for absorption, Lorentz and polarization effects using the Neilson-Howard software package (Howard et al., 1985). The data collection statistics for the native and derivative crystals are presented in Table III-1. Data from the $UO_2(NO_3)_2$ derivative were collected to a higher resolution than the other two derivatives and showed a significant anomalous scattering signal. This anomalous scattering information was included in the calculation of the 5.0 Å MIR phases used in the phased translation search. A Wilson plot (Wilson, 1949) of the native data yielded an upper estimate for the overall temperature factor, $\langle B \rangle = 27$ Å^2.

Table III-1: Data Collection Statistics for Native and Derivative Crystals

	Native	$UO_2(NO_3)_2$	$K_2[Pt(CNS)_6]$	$K[Au(CN)_2$
Measurements	93,839	66,702	26,206	22,267
Unique Reflections	25,084	18,556[1]	5,475	5,631
Resolution (Å)	2.7	3.5	4.5	4.5
Average $I/\sigma I$ (d_{min} Å)	47.2(4.57)	50.2(6.50)	34.7(7.69)	25.2(7.69)
	28.4(3.63)	39.8(5.50)	19.0(6.11)	13.2(6.11)
	12.6(3.17)	37.3(4.50)	16.8(5.34)	11.7(5.34)
	6.8(2.88)	25.5(4.00)	16.7(4.85)	12.4(4.85)
	4.1(2.67)	14.9(3.50)	12.4(4.50)	11.6(4.50)
R_{merge}[2]	6.3	7.3	8.5	9.5
Completeness (%)	98	97	99	99
Completeness (%) $I/\sigma > 2$	93	92	95	96

1 - Bijvoets have not been merged for this derivative.

2 - R_{merge}, defined as $\Sigma(|<I>-I|)/\Sigma(|<I>|)$

Analyses of Heavy Atom Derivatives:

The derivatives were analyzed using Patterson and difference Fourier methods (Blundell and Johnson, 1976). Analysis of the $K_2[Pt(CNS)_6]$ derivative revealed that there was only one site where the heavy atom bound while the $K[Au(CN)_2]$ and $UO_2(NO_3)_2$ derivatives showed one and five sites, respectively. The v=1/2 Harker section for the $K_2[Pt(CNS)_6]$ derivative is illustrated in Figure III-2, and clearly shows the presence of only one heavy atom. The positions and relative occupancies of the heavy atoms are listed in Table III-2 and the refined heavy atom parameters presented in Table III-3.

The heavy atoms tend to cluster to an area near the origin. This was an unexpected result considering that the crystallographic asymmetric unit contains two Fab fragments with a combined M_r of approximately 100,000. This clustering of the heavy atoms can be attributed to the crystal packing and will be addressed in a later section.

The phases obtained from the heavy atom derivatives are poor. Although the quality of the phases, as measured by the figure of merit, is adequate, this result must be tempered by the fact that the phases only extend to low resolution. Moreover, the $K[Au(CN)_2]$ derivative is of extremely poor quality. An initial data set for this derivative was collected from a crystal that had developed a large crack during the heavy atom soak. Analysis of this data set indicated that

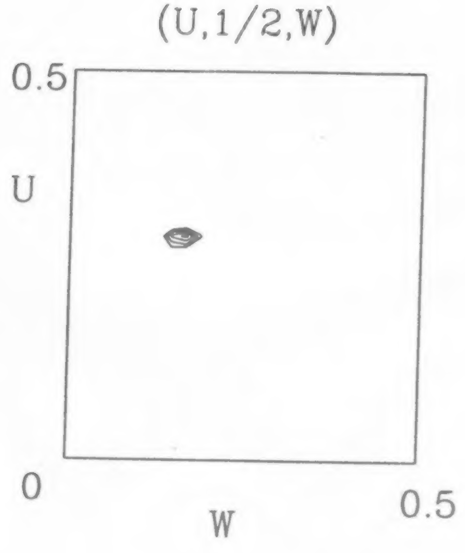


Figure III-2: Harker Section (U,1/2,W) of the Patterson Function for the $K_2[Pt(CNS)_6]$ derivative of Jel 72(I). The section clearly shows the presence of only one site of heavy atom substitution.

Table III-2: Heavy Atom Coordinates and Occupancies

Derivative Site	Coordinates (Å)			Rel. Occ.	Anom. Occ.	Isotropic B-factor	Where Bound
	x	y	z				
$UO2(NO_3)_2$							
1	7.89	5.41	15.92	1.22	0.42	25	F1-GLU H50
2	10.01	16.60	12.31	1.01	0.30	47	F2-GLU H50
3	27.49	28.87	30.42	0.25	0.06	17	F1-ASP H73
4	15.80	91.00	42.97	0.12	0.06	50	F2-GLU H1
5	29.62	88.17	40.12	0.13	0.04	40	F2-GLU L84
$K_2[Pt(CNS)_6]$							
1	15.79	14.92	17.42	0.71	ND	40	F2-HIS L93
$K[Au(CN)_2]$							
1	1.53	18.98	10.15	0.49	ND	40	F1-TYR H106

The occupancies listed are on an arbitrary scale and are not strictly comparable between derivatives.

The residues where the heavy atoms bind are prefaced by F1 or F2 to indicate that they belong to the first or second Fab fragment, respectively.

Table III-3: Refined Heavy Atom Parameters

	Resolution Shell Lower Limit (Å)								
	8.89	8.00	7.27	6.67	6.15	5.71	5.33	5.00	Total
Reflections	213	259	291	364	434	494	584	672	3311
f_m[1]	0.77	0.75	0.74	0.80	0.76	0.71	0.72	0.61	0.72
$UO2(NO_3)_2$									
R_{cullis}[2]	43	44	39	44	40	41	60	47	44
Phasing Power[3]	1.49	1.33	1.15	1.67	1.71	1.31	1.45	1.14	1.34
RMS Anom. Phasing Power [4]	5.75	6.33	1.09	4.15	2.74	2.47	3.35	3.38	2.74
$K_2[Pt(CNS)_6]$									
R_{cullis}	67	59	65	57	67	64	66	86	65
Phasing Power	0.75	0.94	1.03	1.23	1.24	1.17	1.39	0.72	0.98
$K[Au(CN)_2]$									
R_{cullis}	93	83	78	82	78	99	88	87	86
Phasing Power	0.46	0.41	0.48	0.60	0.51	0.60	0.50	0.33	0.46

1 - figure of merit
2 - R_{cullis}, $\Sigma||F_{PH} +/- F_P| - F_{H(calc)}| / \Sigma|F_{PH} - F_P|$
3 - Phasing Power, $<\Sigma F_H> / E_{rms}$
4 - Anomalous Phasing Power, $\Sigma|F_{PH+} - F_{PH-}| / E_{anom}$

there was one site of heavy atom substitution. The data that were used to calculate the MIR phases were collected from another crystal that was not cracked. Analysis of these data indicated that it contained the exact same single site of heavy atom substitution as the data collected from the cracked crystal. This apparent reproducibility was the primary reason for including the data from the $K[Au(CN)_2]$ in the calculation of the MIR phases.

Although some attempts were made at tracing the polypeptide chains of the Fab fragments from the electron density maps calculated with the heavy atom phases, it is unlikely that this approach would have met with much success. The quality of the phases, and the resolution to which they extended, was simply insufficient for this purpose. The phases proved to be more than adequate, however, for use in the phased translation search in the determination of the structure by the molecular replacement method.

Molecular Replacement:

Rotation functions were calculated using the Crowther fast rotation function (Crowther, 1972), which uses a spherical harmonic approximation to decrease the time required for the computation. The Euler angles (alpha, beta, gamma) that define the orientation are also those defined by Crowther (1972). The search models used in these calculations were first superimposed according to their variable light chain domains with the pseudo-twofold of the $V_H \cdot V_L$ domain pair aligned with the X-axis, and the elbow axis positioned parallel to the Z-axis and passing through the point X=Y=0.0. Rotation peaks from different models corresponding to the variable domain pairs would have very similar rotation angles, while those peaks arising from the constant domain pairs should differ mainly in their gamma rotation angle as the elbow angle of the model Fabs varies, because it is the gamma rotation angle that is applied first to the molecule.

The constant domain pair usually gives a clear signal in the rotation and translation searches but the discrimination between the correct peak and the background noise for the variable domain pair varies depending upon which model structure is employed in the calculations. Aligning all of the search models in the above orientation facilitated the interpretation of the results obtained for the variable domain pair. Rotation functions were calculated on **F** and for any particular model domain pair a series of rotation functions were calculated spanning several resolution ranges from 4.0 - 6.0 Å to 4.0 - 12.0 Å, and with the Patterson integration radius varying from 4.0 - 15.0 Å to 4.0 - 24.0 Å. Generally, a wide resolution range, large integration radius and a moderate number (1200-1500) of both model and observed reflections gave the clearest signal. Self-rotation functions were calculated in a similar way.

Self-rotation functions are calculated for crystals which have more than one molecule in the asymmetric unit and could, therefore, possess non-crystallographic symmetry. The kappa = 180° section of the self-rotation function is of particular interest, since it is in this section that the presence of non-crystallographic twofold axes will be apparent. The kappa = 180° section of the self-rotation function for Jel 72 (Figure III-3) indicates that there is a non-crystallographic twofold axis along the diagonal between the crystallographic **a** and **c** axes. The presence of this non-crystallographic twofold axis raises the possibility that the crystals possess pseudosymmetry indicative of a higher order space group, specifically the tetragonal space group $P4_12_12$. If the axes of the crystal are permuted so that the 102.6 Å unit cell dimension becomes the crystallographic **c** axis, then the two remaining cell dimensions (a = 92.4 Å and b = 94.6 Å) have very similar lengths. For the space group to actually be tetragonal these two cell dimensions would have to be identical and there would be a 4_1 screw axis parallel to the

102.6 Å crystallographic axis. We, thus, expected to find a pseudofourfold screw axis parallel to the 102.6 Å cell dimension of the orthorhombic crystals.

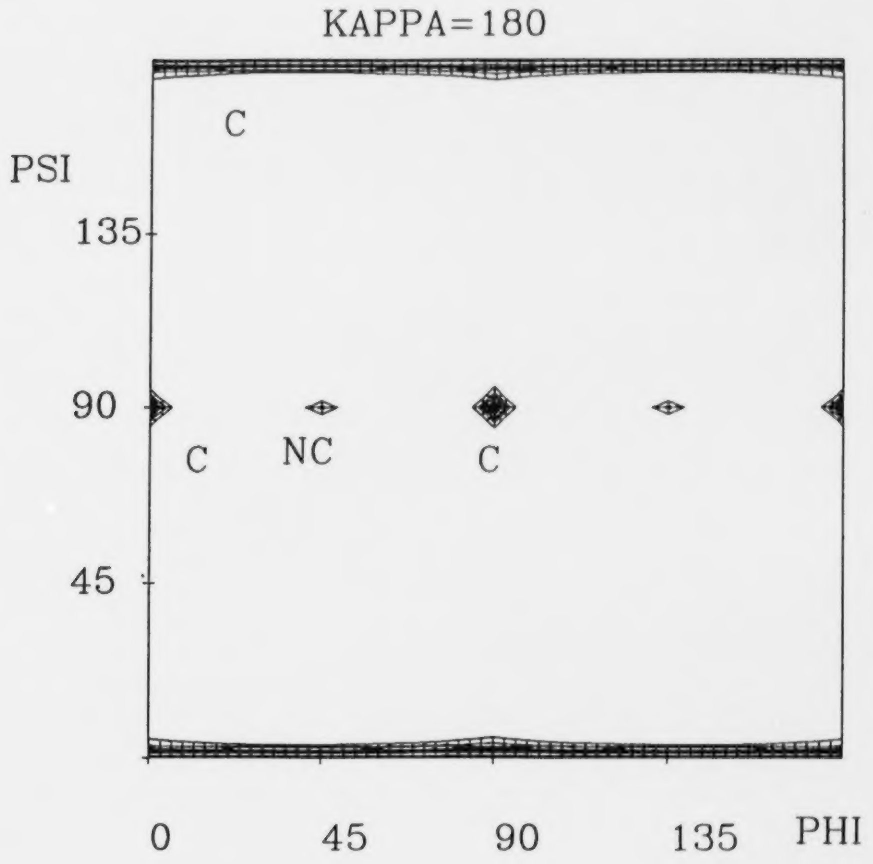


Figure III-3: Section kappa = 180° of the Self-Rotation Function for Jel 72. The three non-intersecting crystallographic twofold axes are indicated by **c**. The non-crystallographic twofold axis is indicated by **nc**.

Rotation functions were calculated using a variety of known Fab domain pairs as models. In this way, the variable domain pair of Fab McPC603 (Satow et al., 1986) was matched with the constant domain pair of Fab Hed10 (Cygler et al., 1987) to give the initial model for molecule number one, while molecule two was derived from the variable domain pair of Hed10 and the constant domain pair of Fab Jel 318. The rotation function results for these four domain pairs are listed in Table III-4.

The oriented domain pairs were placed within the unit cell using a phased translation search (Read and Schierbeek, 1988) employing the 10.0 to 5.0 Å resolution MIR phases. Since the phases from isomorphous replacement may have the incorrect hand two phased translation functions are calculated for each oriented domain pair. The first map is computed with phases of the form (α_p - α_m) while the second map is computed with phases ($-\alpha_p$ - α_m), where α_p are the phases calculated from isomorphous replacement and α_m the phases derived from the oriented domain pair placed arbitrarily within the unit cell. The results from the phased translation searches are presented in Table III-5.

Three of the domain pairs showed clear rotation and phased translation peaks, while the variable domain pair of molecule two was more difficult to place (Figure III-4a-4d). Using a brute force translation search (BRUTE; Fujinaga and Read, 1987), correct translations for the three "good" domain pairs

Table III-4: Rotation Function Results

	Fab-1		Fab-2	
	VH·VL	CH·CL	VH·VL	CH·CL
Model	McPC603	Hed10	Hed10	Jel 318
Resolution (Å)	4 - 10	4 - 10	4 - 10	4 - 10
Patterson Integration Radius (Å)	4 - 24	4 - 24	4 - 24	4 - 24
Reflections Used				
Crystal	1698	1698	1698	1698
Model	2093	1456	1631	1414
α	85.4°	96.3°	83.4°	84.4°
β	13.8°	16.2°	70.7°	73.6°
γ	141.6°	111.1°	320.0°	302.1°
Peak Height[1]	1.31	1.14	0.65	1.29

1 - Peak heights are given as the ratio of the correct peak to the first spurious peak

Table III-5: Phased Translation Function Results

	Fab-1		Fab-2	
	VH·VL	CH·CL	VH·VL	CH·CL
Model	McPC603	Hed10	Hed10	Jel 318
Peak Height[1]				
$(\alpha_p - \alpha_m)$	2.15	2.55	1.04(1.73)[2]	2.22
$(-\alpha_p - \alpha_m)$	1.02	1.41	1.02(ND)	1.01
Translation (Å)				
x	90.0	90.0	86.7(88.4)	88.3
y	50.0	50.0	23.4(25.1)	23.4
z	92.9	93.0	25.4(28.9)	28.8
RMS[3] (Å)	1.19	0.79	5.23	0.70

1 - Peak heights are listed as the ratio of the first to the second peak in the translation search

2 - Numbers in parentheses are those for the model oriented according to the correct rotation angles

3 - RMS, the deviation between the initial placement of the model and the final refined position of the domains

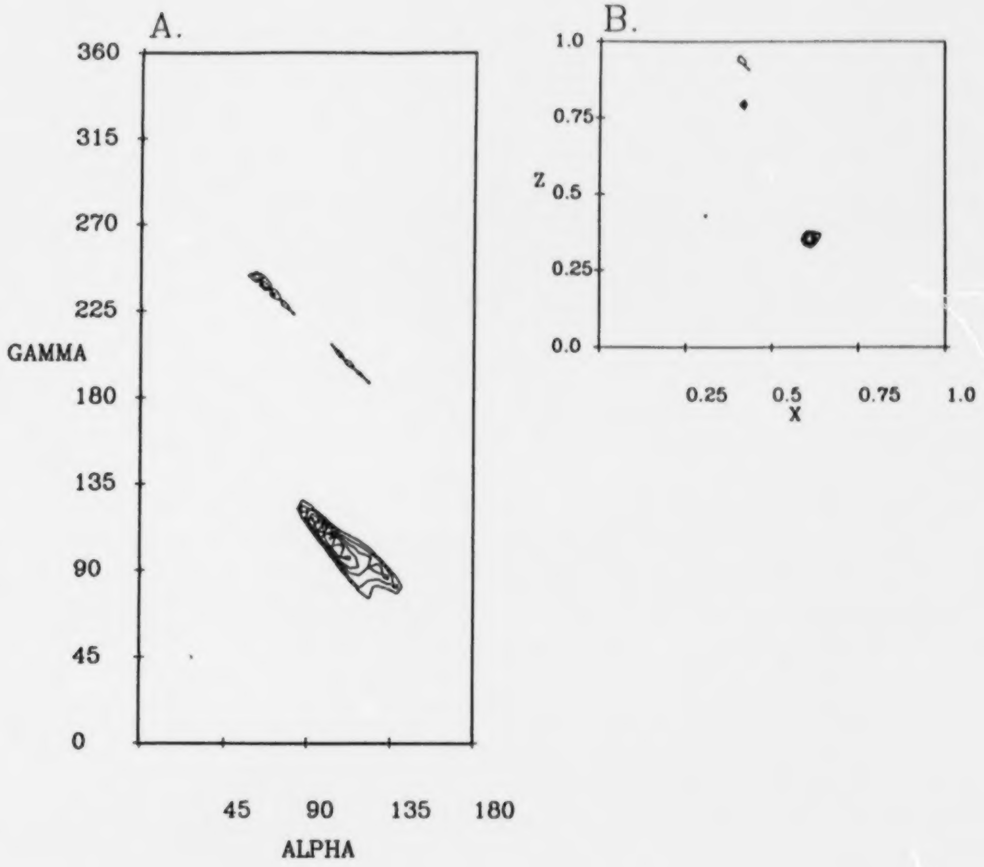


Figure III-4a: Rotation and Phased Translation Searches for Constant Domain Pair One of Jel 72(I) with Hed10 $C_H \cdot C_L$ domains as model. A) Section $\beta = 15^\circ$ of rotation function, B) Section Y=0.54 of phased translation search.

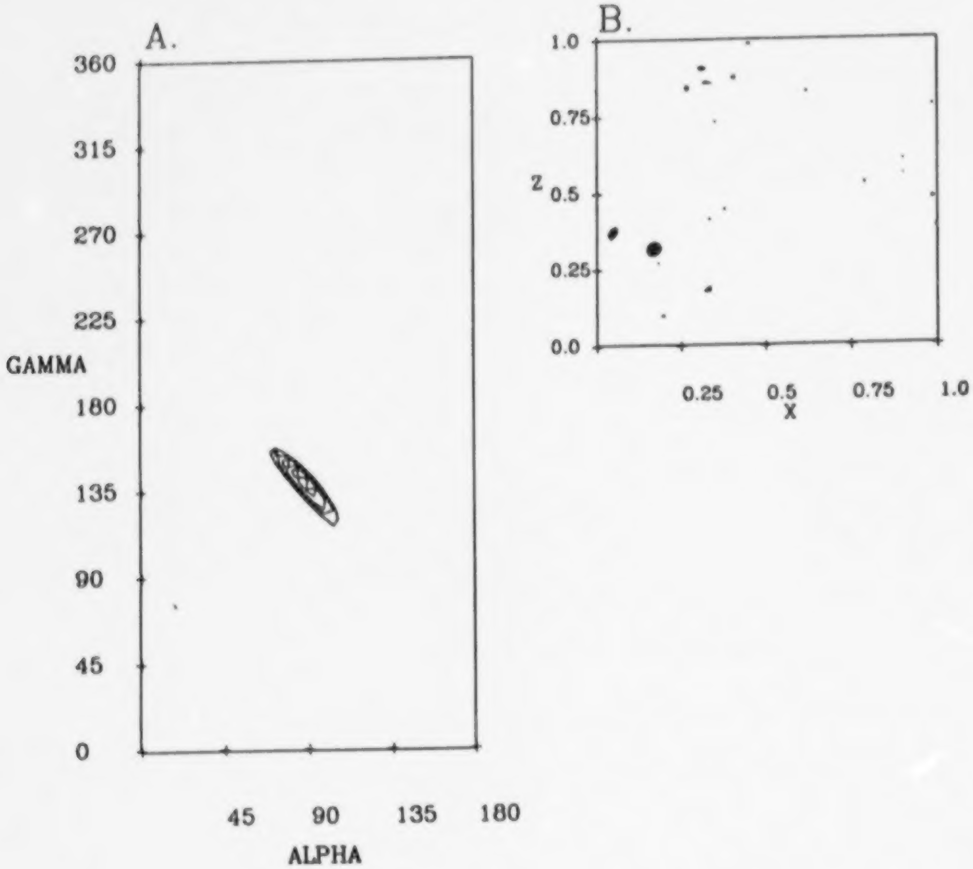


Figure III-4b: Rotation and Phased Translation Searches for Variable Domain Pair One of Jel 72 with McPC603 $V_H \cdot V_L$ domains as model. A) Section $\beta = 15^\circ$ of rotation function, B) Section Y=0.54 of phased translation search.

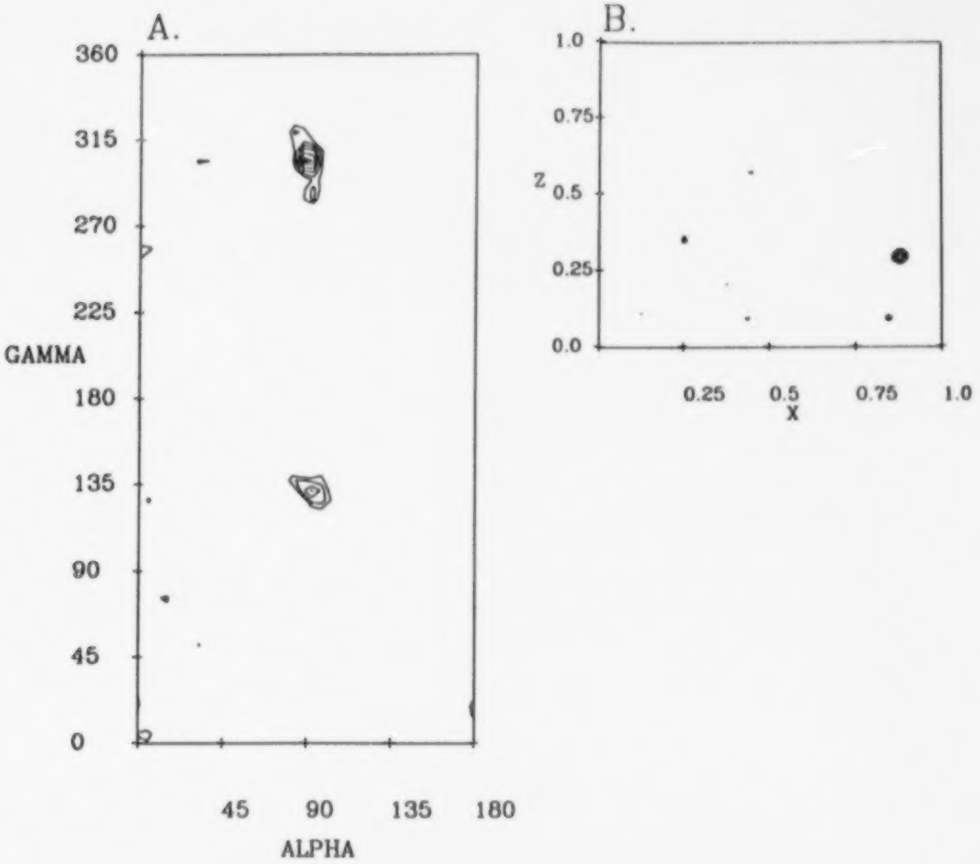


Figure III-4c: Rotation and Phased Translation Searches for Constant Domain Pair Two of Jel 72 with Jel 318 $C_H \cdot C_L$ domains as model. A) Section β = 75° of rotation function, B) Section Y=0.25 of phased translation search.

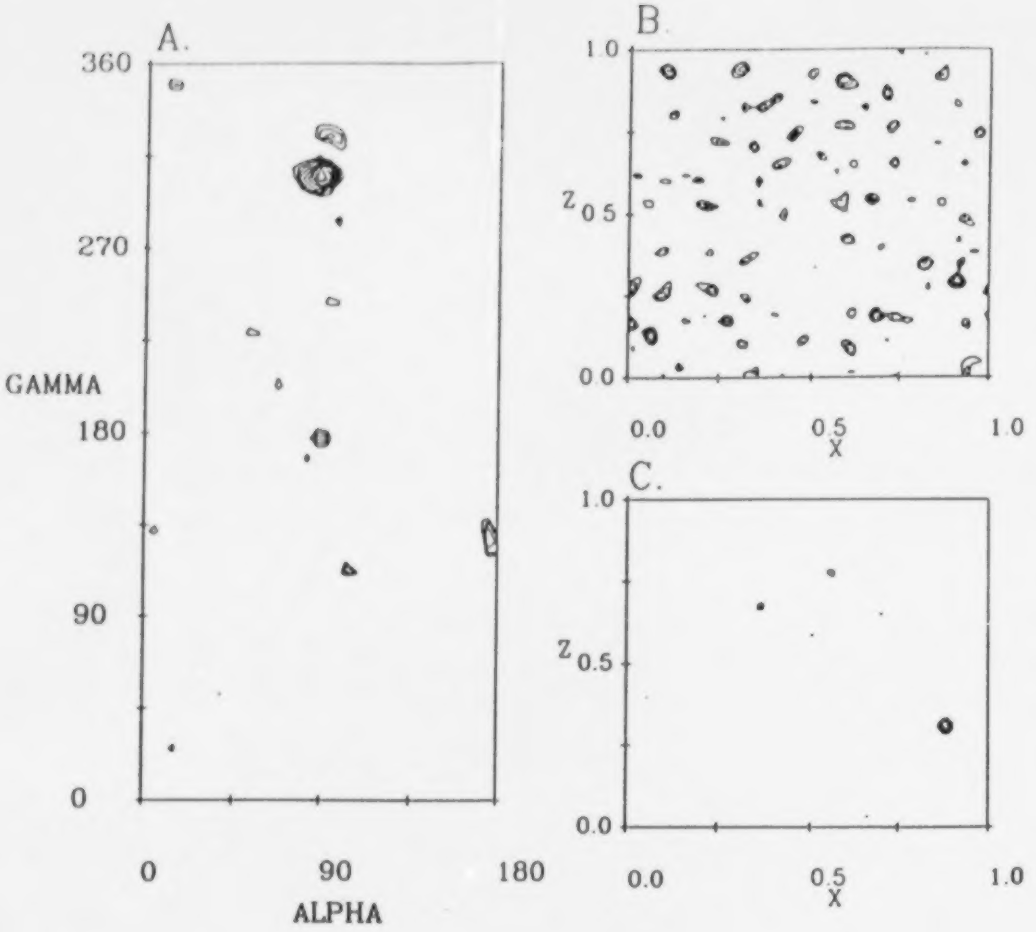


Figure III-4d: Rotation and Phased Translation Searches for Variable Domain Pair Two of Jel 72 with Hed10 $V_H \cdot V_L$ domains as model. A) Section $\beta = 70^\circ$ of rotation function, B) Section Y=0.54 of phased translation search using model oriented according to large rotation peak at gamma=310°, C) Same section of phased translation search using model oriented according to smaller rotation peak at gamma=320°.

could be discerned, but for the variable domain pair of molecule two the correct translation could not be found within the top twenty peaks. Although it still gave a peak in the phased translation search, the signal was not as clear as for the other three domain pairs where the correct peak was more than twice as large as the next highest peak. We interpreted these results as indicating that the model of Hed10 used in the calculations for this variable domain pair was a poor one. The amino acid sequence for the variable region of Hed10 was not known at this time and the structure was at a preliminary stage of refinement. The results could also suggest that the rotation angles for the second variable domain pair were in error which, indeed, turned out to be the case. Fortunately, the positioning of this domain pair was not grossly in error and we were able to determine the correct solution by performing rigid body refinement on the incorrectly placed domain pair while holding the other domains in their correct positions.

Structure Refinement:

Refinement of macromolecular crystal structures is performed in order to improve the agreement between the observed structure factors ($\mathbf{F}_{obs}$) and the structure factors calculated from the atomic model ($\mathbf{F}_{calc}$) that has been fitted to the electron density. The course of this process is followed by a measure of agreement termed the reliability index, **R** , which is defined as:

$$\mathbf{R} = \sum_{hkl} \left| \mathbf{F}_{obs}(hkl) - \mathbf{F}_{calc}(hkl) \right| \Big/ \sum_{hkl} \left| \mathbf{F}_{obs}(hkl) \right|$$

The conventional method for minimizing the **R** value is that of least-squares refinement (Konnert and Hendrickson, 1980) followed by inspection and manual rebuilding of the structure with interactive computer graphics. The disadvantages of least-squares refinement are that the algorithms have a limited radius of convergence and are prone to becoming trapped in local energy minima. These limitations require repeated inspection and manual adjustment of the structure during the course of refinement.

Another method for refining macromolecular crystal structures is employed by the X-PLOR system of programs (Brünger et al., 1989). In this approach molecular dynamics is used to search the conformational space of the molecule. The advantages of this technique are that the structure is less prone to becoming trapped in local energy minima and the refinement has a larger radius of convergence than the least-squares method (Brünger et al., 1987). Thus, there is less need for repeated human intervention and manual rebuilding of the structure.

The oriented and translated Fab domains were refined using molecular dynamics as implemented in the X-PLOR system of programs. The molecules were first divided into eight separate parts, corresponding to the V_H, C_H, V_L and C_L domains of each Fab fragment, and the orientations and positions of the individual domains optimized. Rigid body refinement was performed for forty cycles using data between 10.0 and 5.0 Å resolution, and a further eighty cycles with 10.0 to 4.0 Å data. The **R** value converged to a value of 0.427 from an initial value of 0.507.

Since the amino acid sequence for the variable region of Jel 72 had not yet been determined, the sequences of the V_H and V_L domains of Fab Hed10 were incorporated into the model for the variable domain pairs of both Fab fragments of Jel 72. The residues comprising the complementarity determining regions (CDRs), as defined by Kabat and co-workers (1987), and the elbow peptides between the variable and constant domains, were removed. An initial weight for the diffraction energy term was calculted, W_A = 450,000, and 160 cycles of Powell conjugate gradient energy minimization performed prior to the simulated annealing runs. The structure was then refined by molecular dynamics by artificially heating the structure to 2000 Kelvin and quick cooling to room temperature using a stepsize of 25 Kelvin and data between 10.0 and 3.0 Å resolution. Attempts at employing the non-crystallographic restraints energy term resulted in its value becoming inordinately high and dominating the refinement. For this reason the non-crystallographic restraints were not used.

After two cycles of artificially heating and cooling, the structure was energy minimized and two further simulated annealing cycles performed. The **R** value had dropped from an initial value of 0.427, after the rigid body refinement, to 0.297. Electron density maps were calculated with the coefficients ($2F_{obs}-F_{calc}$, α_{calc}) and converted for use with the INSIGHT graphics program (Biosym Technologies, Inc.) on an Iris Silicon Graphics workstation. The electron density was continuous and interpretable and followed the trace of the polypeptide chains for three of the four domain pairs, but was discontinuous and choppy for the variable domain pair of molecule two. These variable domains had become "twisted" during the first simulated annealing cycle and had assumed backbone conformations that were quite different from those of the variable domain pair of the first molecule. The RMS deviations in the positions of the backbone atoms (N,C,Cα) of the residues comprising the conserved framework regions was 2.2 Å for the V_L domain and 1.5 Å for the V_H domain. Thus, these variable domains were repositioned by holding the other six domains constant and performing rigid body refinement on a model variable domain pair placed in the cell according to the original rotation angles and translation vectors for these domains. After the correct position of this variable domain pair had been determined, the rotation function was reexamined and found to contain a peak corresponding to the correct solution, although this peak was considerably smaller than the peak initially taken to be the correct one (Figure III-3d).

The data presented in Table III-5 indicates that three of the domain pairs were placed within approximately 1.0 Å of their correct positions, but the second variable domain pair was considerably mispositioned. The initial placement of these domains positioned them in such a way that the backbone atoms of the β-strands traced through the side chain density and the entire domains were shifted towards the constant region. The model used for these calculations was a rather crude variable domain pair of Hed10 at an early stage of refinement. The poor quality of this model combined with the fact that it represented only 1/4 of the contents of the asymmetric unit may account for the false peak in the rotation function. Nevertheless, the correct peak did appear in the rotation search and this solution was obtainable from the incorrect placement by performing rigid body refinement on only these domains while holding the other six domains constant.

The original placement of the second variable domain pair was considerably misoriented in the gamma rotation, resulting in these domains being outside the radius of convergence of the rigid body refinement, when all eight rigid bodies were refined simultaneously. With the other domains held constant in their correct positions, the refinement was able to converge on the correct solution for the second variable domain pair even though this required moving the domains over 5 Å from their initial positions.

After one further cycle of molecular dynamics refinement and energy minimization on this structure with the newly positoned variable domains, the **R** value diminished further to 0.281 from 0.297.

With the $V_L \cdot V_H$ and $C_L \cdot C_H$ domain pairs of both crystallographically independent Fab fragments correctly positioned, refinement commenced with the inclusion of data to 2.7 Å resolution. The amino acid sequences of the variable domains were determined by Dr. M. Michelle Berry working in Dr. Lee's laboratory. As the sequences became available the residues comprising the V_L and V_H domain CDRS, as well as the elbow peptides, were built into the electron density of OMIT maps (Bhat, 1988), and $2F_{obs}-F_{calc}$ electron density maps calculated with map coefficients to minimize model bias (Read, 1986). Iterative cycles of molecular dynamics refinement followed by inspection, and manual readjusting of the entire structure with interactive graphics resulted in a decrease in the **R** value to 0.242 for data between 6.0 and 2.7 Å resolution. Individual isotropic temperature factors for the atoms were then refined for 40 cycles followed by energy minimization to improve the stereochemistry. It has been my experience with X-PLOR that the structure can be made to have better stereochemistry if the value of the diffraction energy term, W_A, is lowered to as much as 65% of the value calculated by X-PLOR. This has the effect of forcing the refinement to pay stricter attention to maintaining proper geometry rather than minimizing the **R** value. This is especially important for regions of the molecule associated with solvent exposed loop regions that possess higher temperature factors. The molecular dynamics refinement will attempt to maximize the fit of the residues with the weak electron density found in these regions, with the result being that the bond lengths and angles become distorted. Similar observations about X-PLOR have been reported recently for the refinement of the structure of an Fab fragment in complex with influenza virus neuraminidase (Tulip et al., 1992).

At this point in the refinement, residues with strained backbone geometry were identified by inspecting Ramachandran plots (Ramachandran et al., 1963) and rebuilt if necessary. The trace for a portion of the main chain of the third hypervariable region of the V_H domains from residue Gly H96 to Tyr H100b was rebuilt in both Fab fragments.

This structure, with rebuilt CDR H3 hypervariable loops, was subjected to one further cycle of molecular dynamics refinement and a final cycle of conventional positional refinement followed by refinement of individual isotropic temperature factors for the atoms. The **R** value for this structure was 0.196 for 19,562 reflections in the resolution range 6.0 - 2.7 Å. The discrepancy between the measured F_{obs} and the structure factors calculated from the refined model was examined. Following the procedure of Derewenda and co-workers (1992), a total of 367 reflections were omitted from the data set on the basis of the fact that either the ratio F_{obs}/F_{calc} or F_{calc}/F_{obs} exceeded 2.5. It is not uncommon for data collected from area detectors to possess a few bad reflections that escaped being eliminated during the rejection step following data collection. For Jel 72, these bad reflections amounted to less than 2% of the measured data, and eliminating them from the data set results in lowering the R value to 0.188 for 19,136 reflections with intensities greater than 2.5σ, representing 90% of the observable reflections in the resolution range 6.0 - 2.7 Å. The model is in good overall agreement with the electron density, which is essentially unambiguous for the trace of the main chains and the general positions of the side chains. As can be expected with data at this resolution, the details of the side chain conformations and the torsion angles of some of the peptide oxygen atoms are not well determined. The electron density is, in general, clearer for Fab one than for Fab two, and for the Light chains over the Heavy chains. As the structures of both of the crystallographically independent Fab fragments of Jel 72 conformed to the known structures of immunoglobulin Fab fragments, possessed reasonable stereochemistry, and was in good agreement with the observed data, the refinement was halted at this stage. Although there are some weak features in the final electron density maps could that be interpreted as water molecules, no ordered solvent has been included in the structure because of the relatively low resolution of the data.

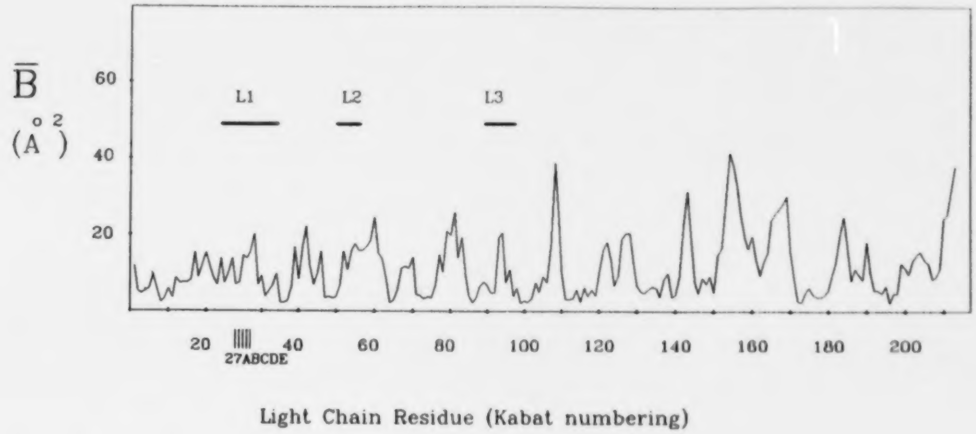
The geometric parameters and refinement statistics for the final structure are presented in Table III-6. A plot of the residue averaged temperature factors is presented in Figure III-5. A representative region of poor quality electron density is illustrated for the tip of the CDR H3 loop of molecule one in Figure III-6. Also shown in the figure is a representative region of good electron density for the H1 loop of Fab one. The density shown is from $2F_{obs}-F_{calc}$ electron density maps contoured at the 1σ level. A Ramachandran plot (Ramachandran et al., 1963) of the main chain torsion angles for all of the residues of the two crystallographically independent molecules is shown in Figure III-7a-7d. The Ramachandran plot shows that the secondary structure of the Heavy and Light chains is primarily β-sheet. There are a few residues that lie outside of the allowed regions on the plot, but they are, for the most part, confined to solvent exposed loops with poor electron density and higher temperature factors (see page 115 forward).

Table III-6: Refinement Statistics for Jel 72(I)

Refinement	Heavy	Light	Total
Residues	2 x (222)	2 x (219)	882
number of non-hydrogen atoms	2 x (1678)	2 x (1706)	6,768

Resolution Range (Å)	Reflections	Shell R-value	Accum. R-value
4.58 - 6.00	2,772	0.1723	0.1723
3.95 - 4.58	2,749	0.1611	0.1666
3.57 - 3.95	2,643	0.1759	0.1694
3.30 - 3.57	2,528	0.1846	0.1725
3.10 - 3.30	2,427	0.2030	0.1769
2.94 - 3.10	2,302	0.2105	0.1805
2.81 - 2.94	2,109	0.2157	0.1844
2.70 - 2.81	1,606	0.2670	0.1885
Total Reflections $I > 2.5\sigma I$	19,136		
Completeness (%)	90		

Geometry	RMS	Violations
Bond Lengths	0.017	59 > 0.060
Bond Angles	3.859	82 > 12
Dihedral Angles	29.00	0 > 90
Deviations from Planarity		
Improper Angles	1.55	0 > 20
Peptide Bond	2.26	0 > 15



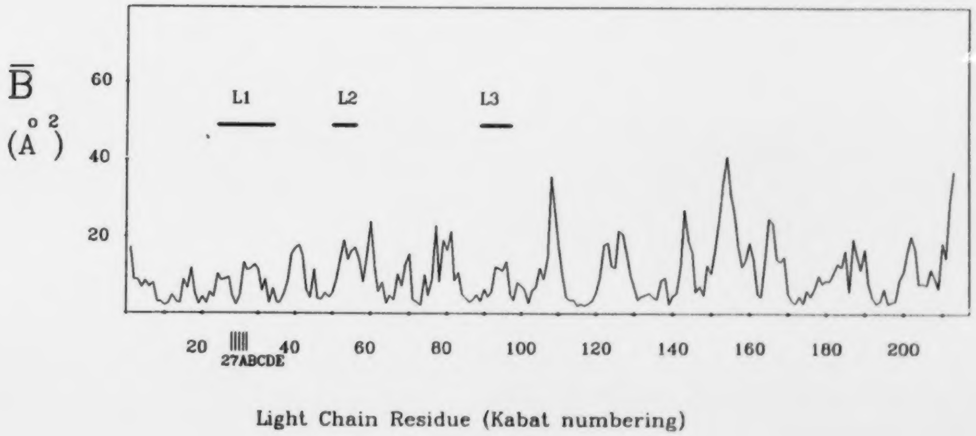


Figure III-5a: Temperature Factor Distribution averaged over all atoms for the Light chains of Jel 72(I). Light Chain of Fab-1 (top), and Light Chain of Fab-2 (bottom). The positions of the CDRs are indicated. Residue numbering follows that of Kabat and co-workers (1987) with the positions of insertions as indicated and deletions labelled in parentheses.

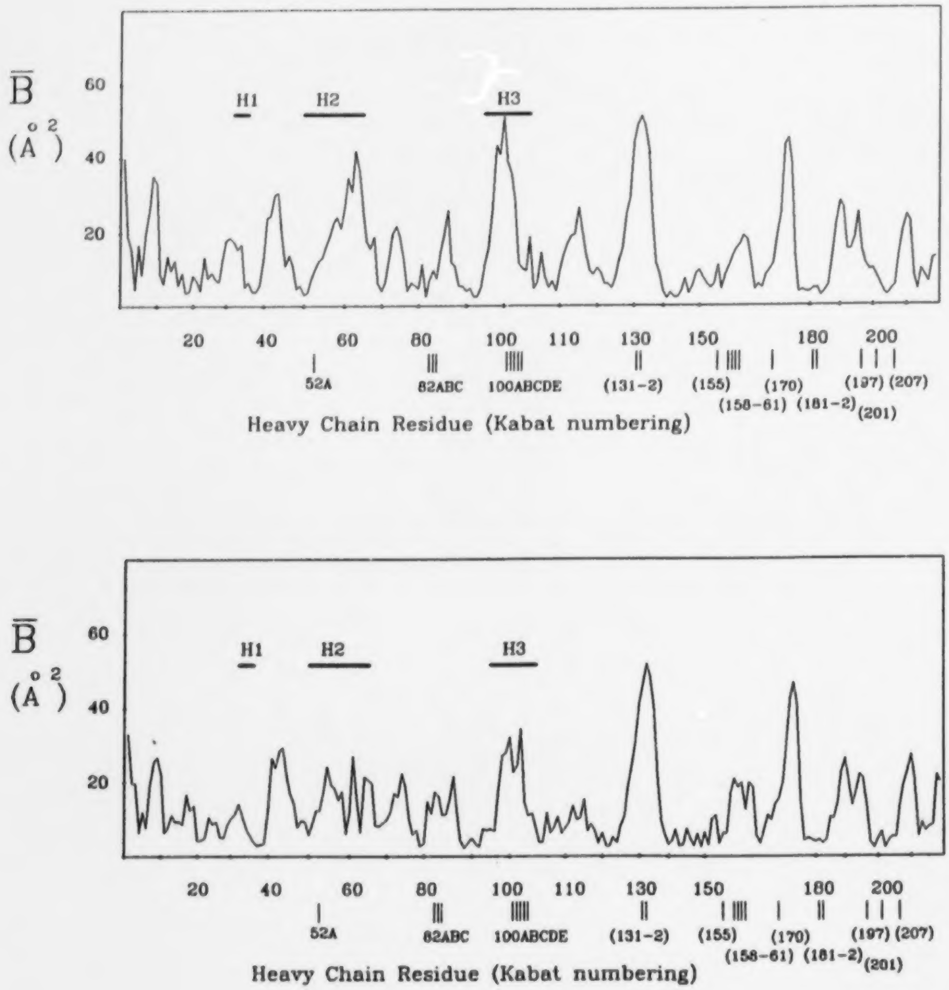


Figure III-5b: Temperature Factor Distribution averaged over all atoms for the Heavy chains of Jel 72(I). Heavy Chain of Fab-1 (top), and Heavy Chain of Fab-2 (bottom). The positions of the CDRs are indicated. Residue numbering follows that of Kabat and co-workers (1987) with the positions of insertions as indicated and deletions labelled in parentheses.

Figure III-6: Regions of good and poor quality electron density for Jel 72(I). Portion of $2F_{obs}-F_{calc}$ electron density maps contoured at the 1σ level showing good density for CDR H1 and CDR H2 of Fab one (top), and poor quality density for the residues near the interchain disulphide in the C_H domain of Fab one (bottom).

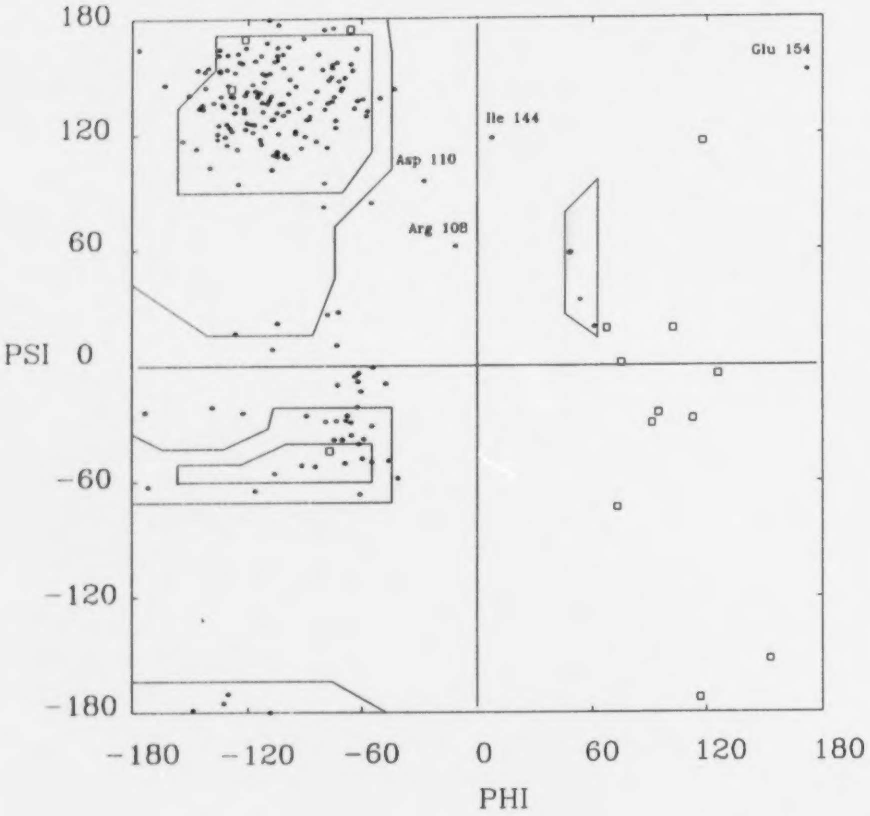


Figure III-7a: Ramachandran Plot of the main chain torsion angles for the Light chain of Fab one of Jel 72(I). The positions of glycine residues are indicated by the boxes, all other residues by black circles. The residues with strained backbone geometry (see text) are indicated by their three letter amino acid code and their Kabat residue number.

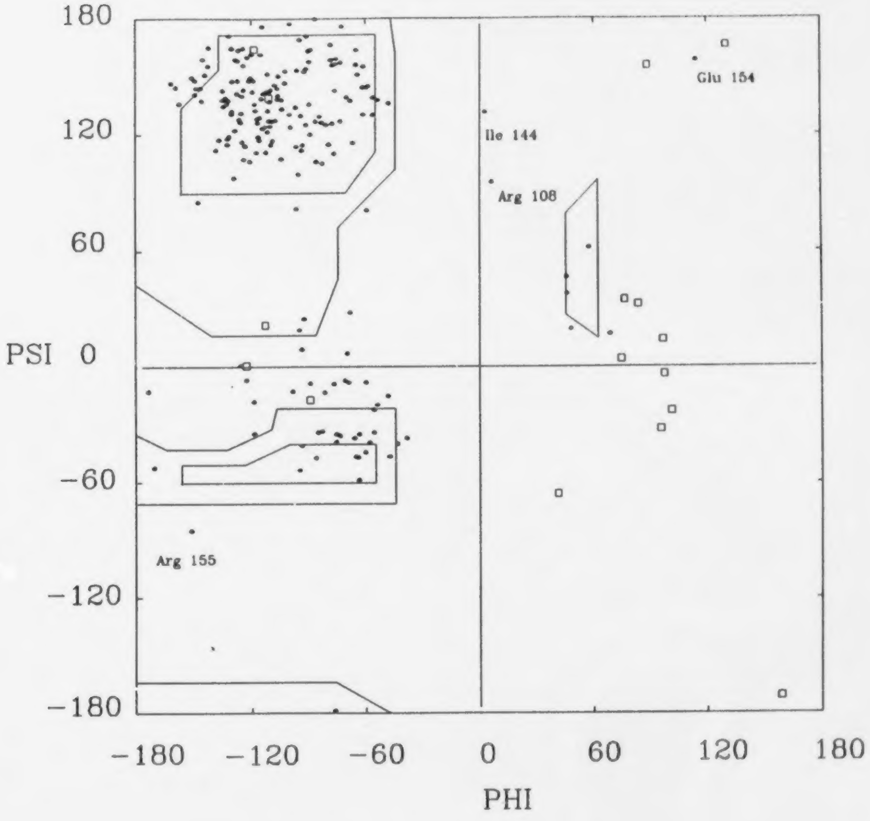


Figure III-7b: Ramachandran Plot of the main chain torsion angles for the Light chain of Fab two of Jel 72(I). The positions of glycine residues are indicated by the boxes, all other residues by black circles. The residues with strained backbone geometry (see text) are indicated by their three letter amino acid code and their Kabat residue number.

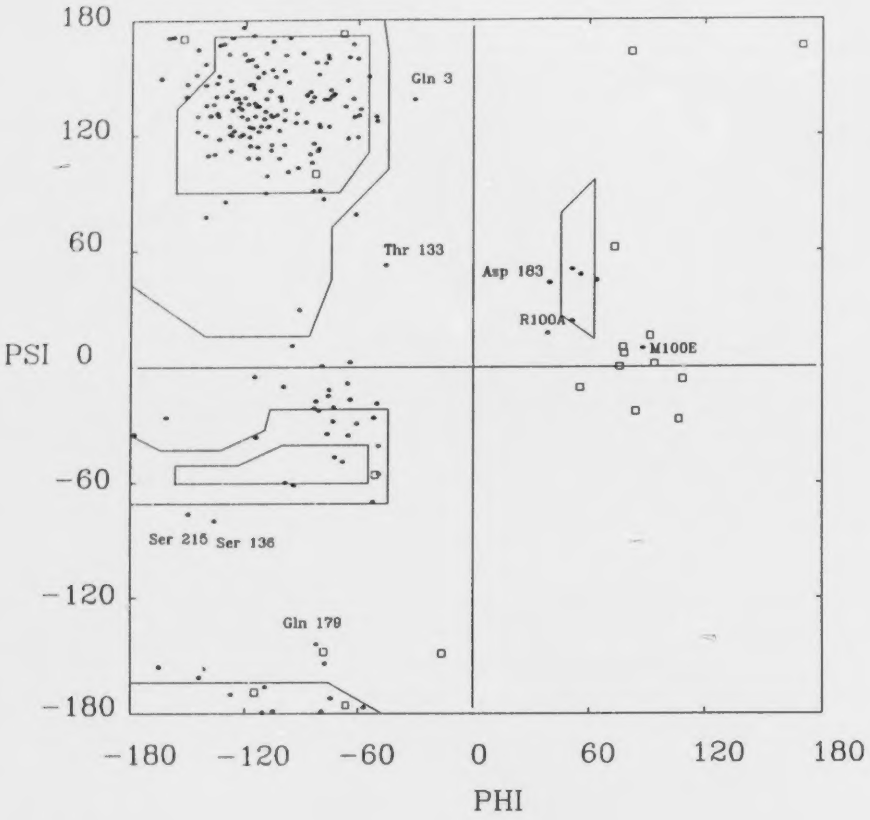


Figure III-7c: Ramachandran Plot of the main chain torsion angles for the Heavy chain of Fab one of Jel 72(I). The positions of glycine residues are indicated by the boxes, all other residues by black circles. The residues with strained backbone geometry (see text) are indicated by their three letter amino acid code and their Kabat residue number.

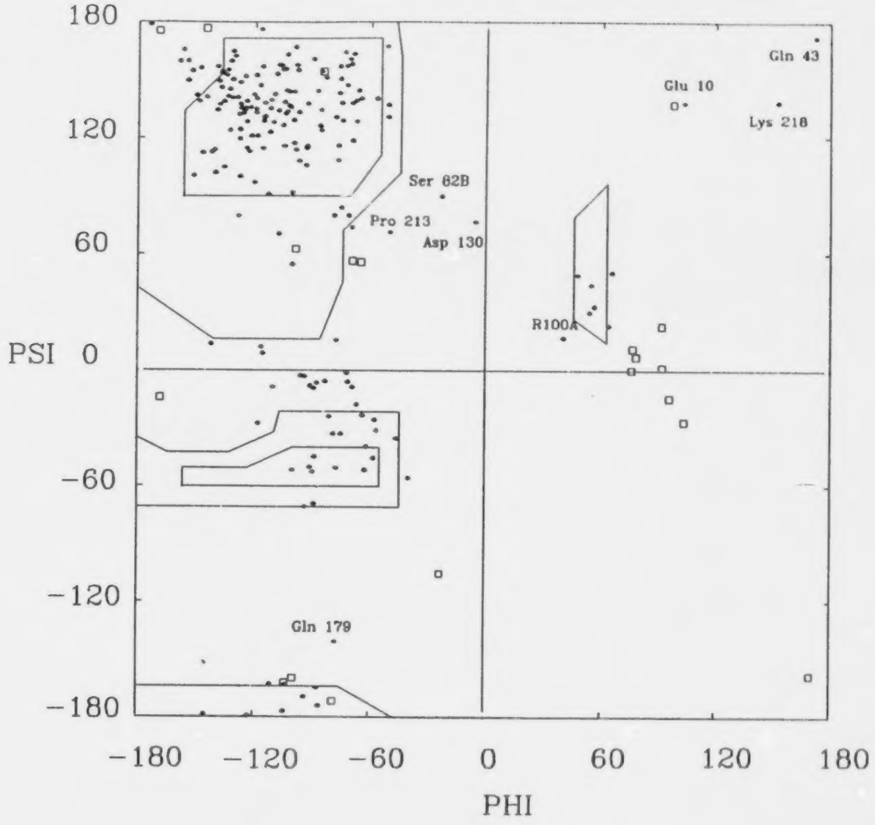


Figure III-7d: Ramachandran Plot of the main chain torsion angles for the Heavy chain of Fab two of Jel 72(I). The positions of glycine residues are indicated by the boxes, all other residues by black circles. The residues with strained backbone geometry (see text) are indicated by their three letter amino acid code and their Kabat residue number.

Crystal Packing:

The packing of the Fabs in the unit cell is unusual. Fab fragments often pack tightly in a "head-to-tail" fashion, that is, their variable regions pack against the constant regions of symmetry related Fab fragments (Cygler et al., 1987). Using this nomenclature, Jel 72 can best be described as packing in a "head-to-head" manner, with two variable regions related by non-crystallographic symmetry packing against one another. The packing of the Fabs within the unit cell is illustrated in Figure III-8. The packing is tight, with no room between the variable regions to accommodate a double stranded DNA helix. Another consequence of the packing is the clustering of heavy atom sites to the area between the two variable regions. The $Pt(CNS)_6^{2-}$ derivative binds to histidine L93 of Fab-1 in between the two variable regions rather than to an expected site on the constant heavy chain as has been reported for two previous Fab structures (Padlan et al., 1973; Navia et al., 1979).

The heavy chains of the non-crystallographically related constant domains of Jel 72 pack against one another, occluding what would have been the $Pt(CNS)_6^{2-}$ site. The platinum derivative could only be obtained if the crystals were soaked in extremely dilute solutions of the heavy atom. At any higher concentrations the crystals would shatter.

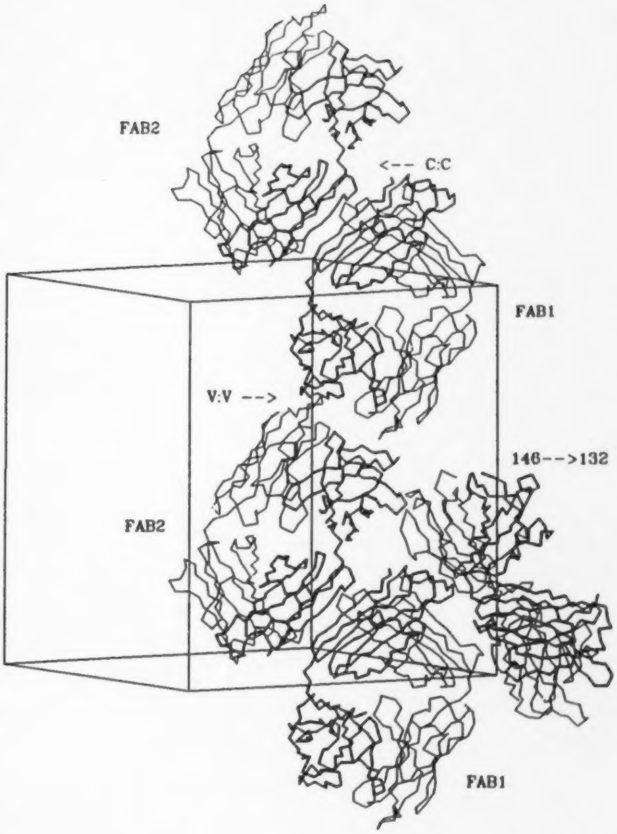


Figure III-8: Crystal Packing of Jel 72(I). For each Fab fragment the Heavy chain is drawn in bold lines, the Light chain in thin lines. The head-to-head packing of the variable domains and the packing of the C_H domains is indicated. The direction of the elbow bend from 146° to 132° for a symmetry related molecule of Fab two (F2R) is also indicated. All of the Fab fragments displayed lie in approximately the plane of the page.

The residues that contact symmetry related molecules within a radius of 4.0 Å are listed in Table III-7. While the crystal contacts made by the constant region of Fab-2 are predominantly through this Heavy chain interaction, the constant region of Fab-1 also makes close contacts with the Light chain of another replicate of Fab-2. Close contacts are made by both the Light and Heavy constant chains of Fab-1 one with the variable and constant Light chains, as well as the elbow region, of this replicate of Fab-2. As a consequence of this type of packing, the elbow angles of the two Fabs are different. Fab-2 has an elbow angle of 132°, while Fab-1 possesses an elbow angle of 146°. In order for the constant region of Fab-1 to fit in between the two symmetry related constant regions of Fab-2, as shown in Figure III-8, the elbow angle of Fab-2 must be compressed. If Fab-2 were to have the more extended elbow angle of 146° then there would simply not be enough room for the constant region of Fab-1 to pack in the crystal.

As expected from the self-rotation function results, the crystals possess pseudo-symmetry indicative of the tetragonal space group $P4_12_12$ (Figure III-9), but because the packing interactions dictate that the two Fabs have different elbow angles, they can not adhere to strict tetragonal symmetry. Instead, the two molecules are related by a pseudo-fourfold screw axis parallel to the Y axis of the crystal.

Table III-7: Crystal Contacts for Jel 72(I)

L1:	L2:	Symmetry Mate
3,5,8-10,12	14-15,108,110,199-202	4
197,199-203,213	7-12,145,147,213	2
	H2:	
18,74,144-147,154-156, 161,163,175,197	65,82B,128-134, 199-200,223B-223C*	4
27D-27E,93,123	**33,52,54,56-58,64**	3
	H1:	
184,187-188	19,72,75,77,79	2
H1:	H2:	
1-3,101-102, 164-165	**1,26-27,100**	1
135,199	177-178	2
214-222	215-222	3
L2:	H1:	
144-146,155-157, 160,161,163,175	125-135,196,199-202, 223B-223C*	2
27-27A,27C-27E,93	**30,33,53-56**	3
	H2:	
184-185,188	19,70-73,79-81	4

* - denotes the C-terminal residue of the Heavy Chain. Residues within the CDRs are listed in bold type.

Symmetry Mate:
1 - (x, y, z)
2 - (1/2-x, -y, 1/2+z)
3 - (1/2+x, 1/2-y, -z)
4 - (-x, 1/2+y, 1/2-z)

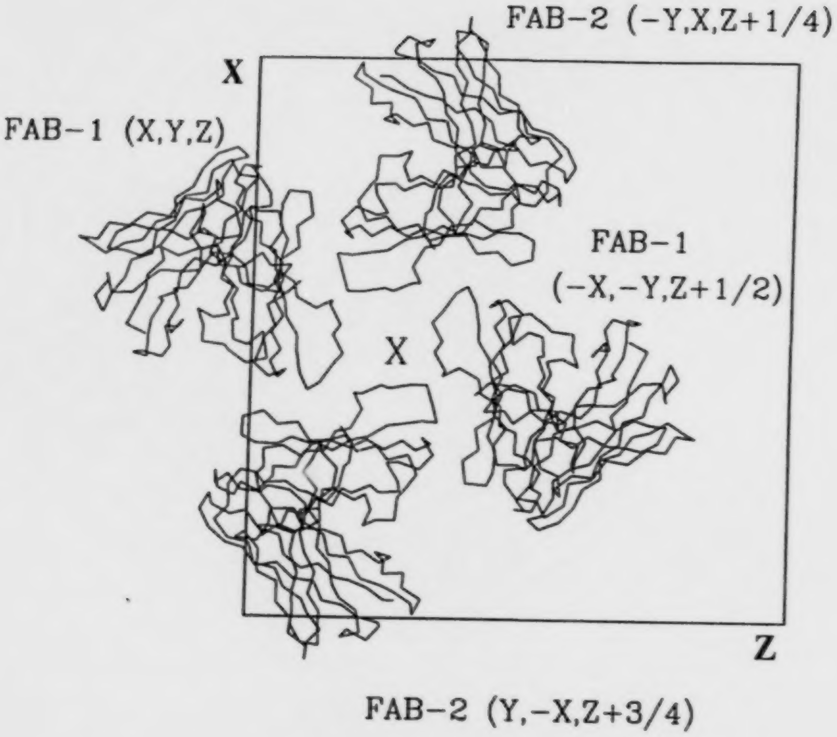

FAB-2 (-Y,X,Z+1/4)
X
FAB-1 (X,Y,Z)
FAB-1
(-X,-Y,Z+1/2)
X
Z
FAB-2 (Y,-X,Z+3/4)

Figure III-9: Pseudo 4_1 symmetry of Jel 72(I). View down the Pseudo 4_1 axis, parallel to the crystallographic **b** axis, of the four symmetry related molecules. Only the Cα atoms of the Heavy chains are shown for clarity. The symmetry operation for each molecule if it were to be in space group $P4_12_12$ is indicated. The approximate position of the pseudo-4_1 axis is marked by **X**.

Although two major isoelectric species were produced during the papain digestion of Jel 72 IgG, suitable crystals of the Fab fragments could be grown from preparations of Fabs that contained only the protein with the higher isoelectric point. Even small amounts of contamination with the lower isoelectric species resulted in crystals that were badly intergrown, with still higher amounts of this species present only microcrystals could be obtained (Boodhoo et al., 1988). No crystals could be grown from pure preparations of the Fab species with the lower isoelectric point under the same conditions that gave crystals with the pure pI 8.8 species. One possible reason for this would appear to be that the C-terminal arginine residues of both heavy chains are involved in several crystal contacts. Presumably the Fab species with the lower isoelectric point had either more or fewer residues at the C-terminus of its heavy chain as a result of the papain digestion. This slight difference in C-termini may have been enough to hamper crystallization at low concentrations and preclude it altogether at higher concentrations.

Structure Quality:

Although the two crystallographically independent Fab fragments of Jel 72 differ significantly in their quaternary structure, as described by the elbow angle that relates the $C_H \cdot C_L$ domain pair to the $V_H \cdot V_L$ domain pair, they are otherwise very similar. A detailed comparison of the two structures is presented in the next section. The structures of both of the Fab fragments of Jel 72 determined in this crystal form conform to the structures that have been previously observed for immunoglobulin Fab fragments. Their elbow angles and the values for their pseudo-twofold rotations, relating the Heavy to Light chain domains, fall within the range observed for other Fab fragments (Cygler and Anderson, 1988).

The Ramachandran plots of the main chain torsion angles for the Heavy and Light chains of Jel 72 show that there are some residues that lie outside of the allowed regions of the plot. These residues are found mostly in solvent exposed loop regions of the constant domains. There is generally poor electron density for these residues and they possess higher than average temperature factors. Consequently, the model for these regions of the molecules is not well determined. Of the residues within the antibody CDRs, only two (arginine H100a and methionine H100e) possess main chain conformations that fall outside of the allowed regions (Figure III-7c-7d). This is not an uncommon occurence in Fab fragment structures (see eg. Strong et al., 1991 and Tulip et al., 1992). Other residues with strained backbone conformations are in regions associated with poor electron density such as the switch region between between the C_L and V_L domains or the COOH terminus of the Light chains.

The electron density for the residues of the antibody CDRs is essentially unambiguous for the trace of the main chains and the general positions of the side chains (Figure III-6). As can be expected with data at this resolution, the torsion angles of some of the peptide oxygens, as well as the details of some of the side chain conformations, are not well determined. The electron density is, in general, clearer for Fab one than for Fab two, and for the Light chains over the Heavy chains. This fact is reflected in the temperature factor distributions for the two Fab fragments (Figure III-5). As has been noted for at least three other immunoglobulin Fab fragment structures (Stanfield et al., 1990; Tulip et al., 1992; and Tormo et al., 1992), the electron density is poor for the COOH terminus of the Light chains (residues L211-L214) and for the region of the Heavy chains (H128-H134) to which it is covalently linked via the interchain disulphide bind between cysteine L214 and cysteine H128. This disulphide is exposed to solvent on the surface of the constant domains and may be in more than one conformation. The mean coordinate error for the structure described here is 0.25 - 0.30 Å (Figure III-10), as estimated by the method of Luzzati (Luzzati, 1952).

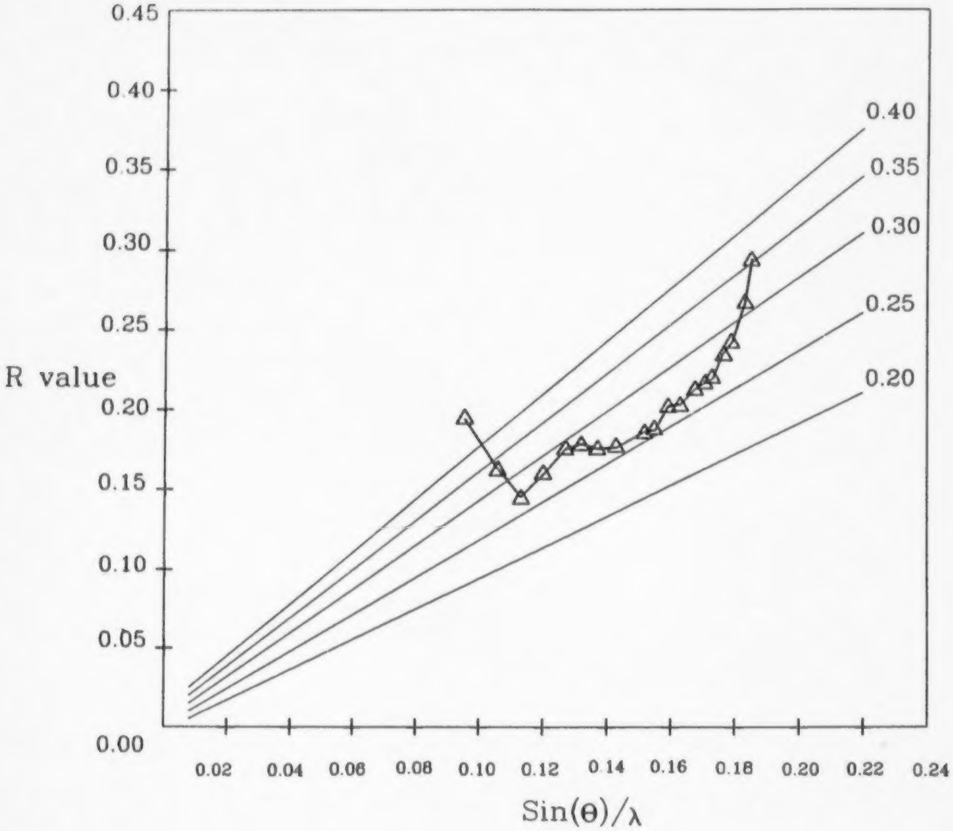


Figure III-10: Plot of R value (triangles) by shell on reflections for which I > 2.5σ(I), versus sin(θ)/lambda , along with the theoretical curves (Luzatti, 1952) for the R value if all discrepancy is attributable to the coordinate error.

Comparison of the two Fabs in the Asymmetric Unit:

The initial misplacement of the variable domain pair of Fab two frustrated attempts to incorporate a non-crystallographic energy restraint term in the molecular dynamics refinement. The restraints used by X-PLOR employ an energy term that increases if the refinement tries to move an atom away from the average position calculated from the positions of the two atoms related by the non-crystallographic symmetry. One of the first indications that the domains were not positioned correctly was a large increase in this energy term at an early stage of refinement. After the variable domain pair of Fab two was positioned correctly, and in all of the subsequent refinement, the non-crystallographic symmetry restraints term was not employed. The two Fab fragments in the asymmetric unit were, thus, refined independently of one another. It is, therefore, not surprising that there are some small differences in their quaternary and tertiary structures.

Although the pseudo-twofold rotations relating the variable and constant domains of each Fab fragment are not identical, they do fall in the range observed for other Fab fragment structures (Cygler and Anderson, 1988). The pseudo-twofold relating the V_H to the V_L domain of Fab one is 173.8°, compared to a rotation of 174.3° observed for the variable domains of Fab two. The C_H and C_L domains are related by a smaller pseudo-twofold angle than the variable domains and are also slightly different from each other. The C_H and C_L domains of Fab one possess a pseudo-twofold angle of 166.8°, while the C_H and C_L domains of Fab two are related by a rotation of 165.5°. These angles were calculated by comparing only those Cα atoms that are considered to be part of the conserved β-sheet core of each domain (Getzoff et al., 1988). The small differences in the quaternary strucuture may indicate that the domain association is flexible and can adapt to the structure of the antigen during binding (Colman, 1988). The differences are small enough, however, that they may simply reflect the inherent error in the coordinates, estimated by the method of Luzzati (Luzzati, 1952) to be 0.25 to 0.30 Å, in which case the real difference in the pseudo-twofold angles may be negligible.

Since the non-crystallographic symmetry restraints were not applied, the accuracy of the model can be evaluated by a comparison of the RMS deviations between the coordinates of the two Fab fragments. The RMS deviation between the Cα atoms of the framework residues is 0.44 Å for the V_L domains, and 0.68 Å for the V_H domains. For the constant domains, the RMS deviations between Cα atoms are 0.43 Å for C_L and 0.78 Å for C_H. These values compare very favourably with those determined from the structure of Fab 131 (Garcia et al., 1992), which also contained two Fab fragments in the asymmetric unit that were independently refined.

The differences in the tertiary conformations of the residues in the two independent Fab fragments can be examined by determining the RMS deviations of the atoms after alignment of the appropriate domains (Figure III-11). These RMS deviations were calculated using X-PLOR for the Cα atoms of each domain. The large deviations seen for the H2 and H3 hypervariable loop reflects the fact that the electron density for these regions in Fab two is weak and ambiguous. The variation in the positions of the Cα atoms is similar to the temperature factor distributions. This correlation may indicate that where the main chain conformations of the two Fab fragments differ is simply a reflection of the positional uncertainty of the coordinates.

When this type of analysis is extended to all of the atoms of the hypervariable loops, a similar conclusion can be reached. The RMS deviations from alignment for the main chain atoms, and for all of the atoms, of the hypervariable loops are listed in Table III-8, and depicted graphically in Figure III-12. The hypervariable loops that have lower temperature factors and, therefore, more accurately determined coordinates are in similar conformations.

Any description of an antibody combining region based on medium resolution structures of the hypervariable loops is likely to be in error at the level of precise atomic detail. For this reason, the structures that we have determined will be described with the emphasis on the shape and chemical nature of the combining site and the relative disposition of the hypervariable loops with respect to each other.

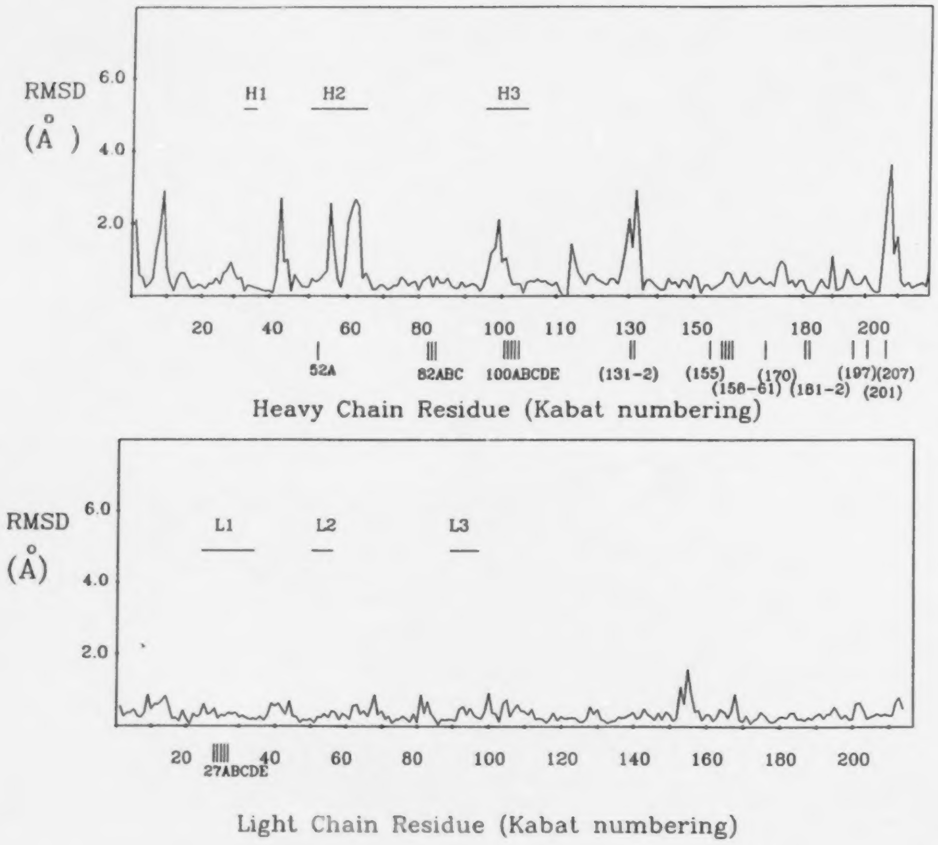


Figure III-11: RMS Differences between Cα positions for the Heavy and Light Chains of the Two Crystallographically Independent Fab Fragments.

Table III-8: RMS Deviations from alignment of the Variable and Constant Domains and Hypervariable Loops of the two Fab Fragments in crystal form I of Jel 72

Loop	Atoms	Backbone	Atoms	All Atoms
C_H	303	0.748	---	----
V_H	354	0.781	---	----
H1	21	0.395	59	0.972
H2	15	0.654	33	0.868
H3	33	0.318	95	1.144
C_L	300	0.380	---	-----
V_L	330	0.403	---	-----
L1	36	0.196	91	0.556
L2	9	0.072	21	0.182
L3	18	0.274	48	1.764

A.

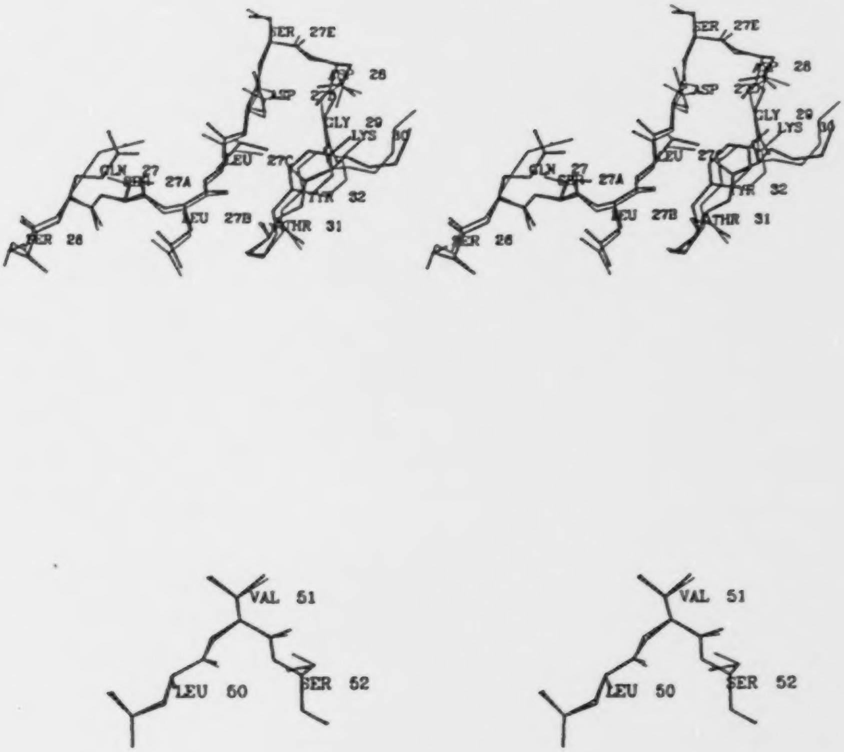

SER 27E
ASP 28
ASP 27D
GLY 29
LYS 30
GLN 27
SER 27A
LEU 27C
TYR 32
LEU 27B
THR 31
SER 26
VAL 51
LEU 50
SER 52

Figure III-12: Hypervariable Loop conformations. Stereoscopic views of aligned hypervariable loops observed in Fab tow (thick lines) and Fab one (thin lines). The loops have been superimposed according to a least squares fit of their backbone atoms (N,C,Cα). **A)** L1 and L2.

B.

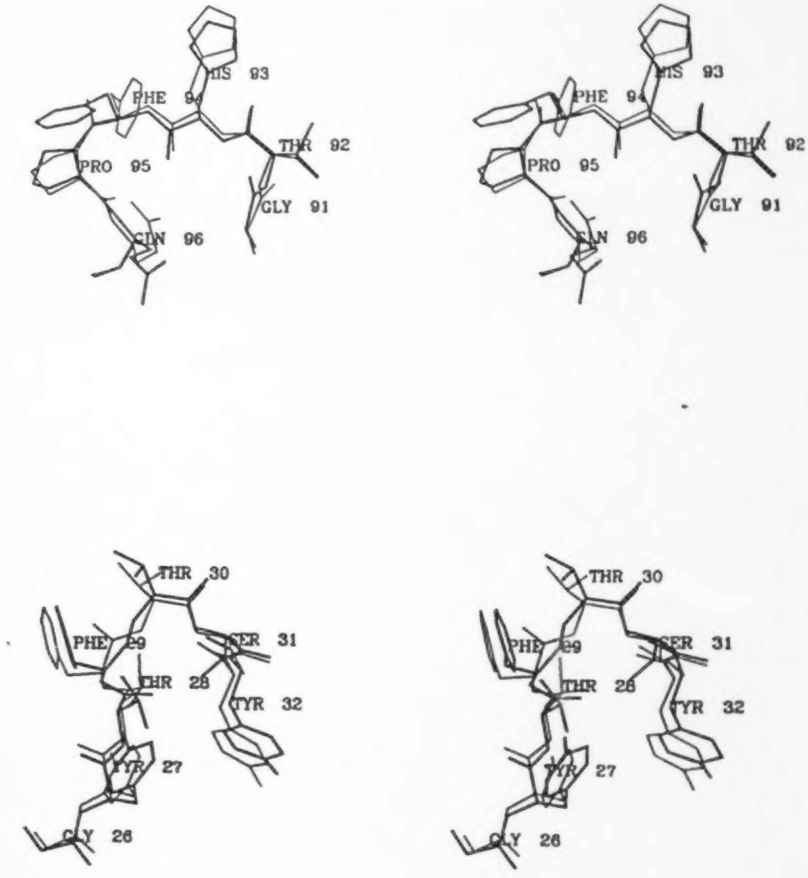

HIS 93
PHE 94
THR 92
PRO 95
GLY 91
GLN 96
HIS 93
PHE 94
THR 92
PRO 95
GLY 91
GLN 96
THR 30
PHE 29
SER 31
THR 28
TYR 32
TYR 27
GLY 26
THR 30
PHE 29
SER 31
THR 28
TYR 32
TYR 27
GLY 26

Figure III-12: Hypervariable Loop conformations. Stereoscopic views of aligned hypervariable loops observed in Fab tow (thick lines) and Fab one (thin lines). The loops have been superimposed according to a least squares fit of their backbone atoms (N,C,Cα). B) L3 and H1.

c.

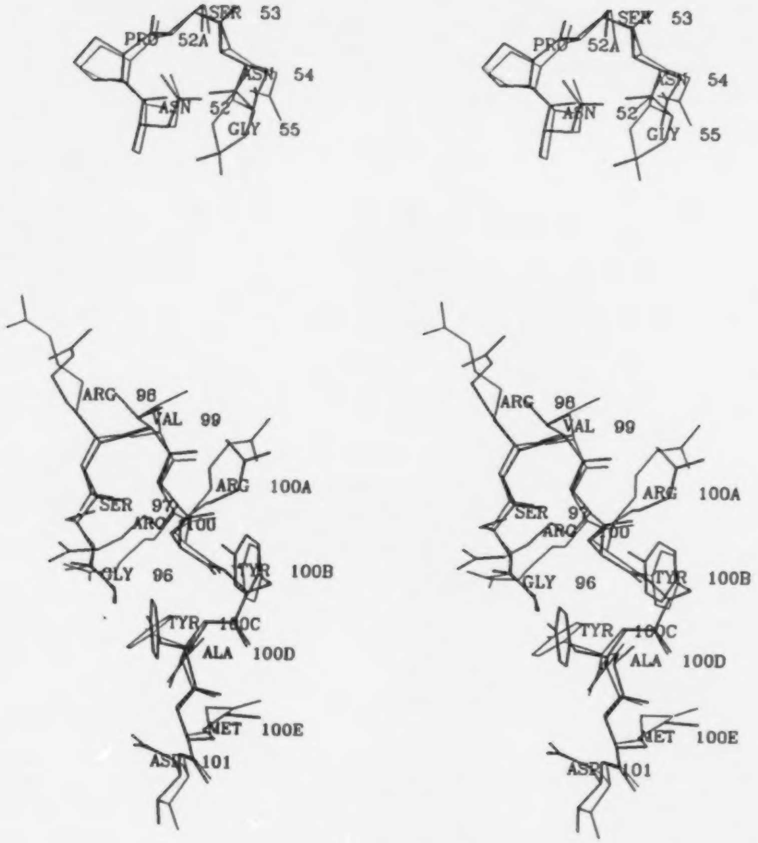

PRO 52A
SER 53
ASN 54
ASN 52
GLY 55
ARG 98
VAL 99
ARG 100A
SER 97
ARG 100
TYR 100B
GLY 96
TYR 100C
ALA 100D
MET 100E
ASP 101

Figure III-12: Hypervariable Loop conformations. Stereoscopic views of aligned hypervariable loops observed in Fab tow (thick lines) and Fab one (thin lines). The loops have been superimposed according to a least squares fit of their backbone atoms (N,C,Cα). **C**) H2 and H3.

The Antibody Combining Region:

The structure of Fab Jel 72 contains the immunoglobulin fold that has been observed in all previously determined antibody Fab fragment structures (Figure III-13a). The antibody combining region of Jel 72 differs from the combining regions observed in antibodies that are directed against single stranded nucleic acids (Herron et al., 1991; Cygler et al., 1987). The antibodies that bind single stranded DNA, as typified by Hed10 (Figure III-13b), have combining sites that form deep clefts, approximately parallel to the $V_H \cdot V_L$ domain interface, that are formed by having long CDR L1 and CDR H2 hypervaraible loops, reminiscent of the combining regions of antibodies that bind small molecular weight haptens such as McPC603 (Padlan et al., 1973). The combining region cleft of Hed10 and BV04-01 is particularly enriched in aromatic amino acids that interact with the DNA bases.

For antibodies that bind double stranded DNA such intimate interactions with the DNA bases can not take place because the bases are involved in Watson-Crick hydrogen bonding interactions in the center of the DNA helix. It is not surprising, therefore, that the combining region of Jel 72 does not possess a pronounced cleft, but instead presents a relatively uniform surface that is formed by having long CDR L1 and CDR H3 loops and a CDR H2 loop that is two residues shorter than the H2 loop observed in Hed10 (Figure III-13). The resulting combining region is rich in aromatic amino acids

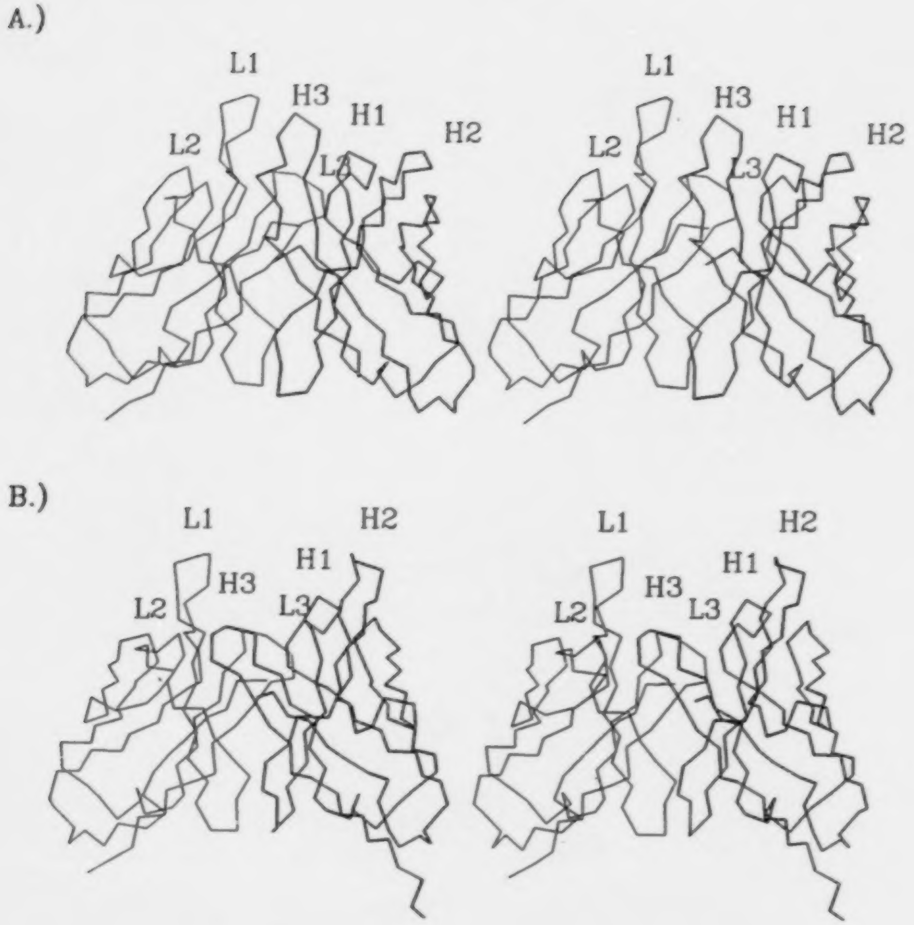

A.)
L1
H3
H1
L2
H2
L3
L1
H3
H1
L2
H2
L3
B.)
L1
H2
H1
H3
L2
L3
L1
H2
H1
H3
L2
L3

Figure III-13: Stereo views of the main chain conformations of the polypeptide backbone for the V_H (thick lines) and V_L (thin lines) domains of, **A**) the double stranded DNA specific antibody Jel 72 and, **B**) the single stranded DNA specific antibody Hed10. The positions of the six hypervariable loops are indicated.

such as tryptophan and tyrosine, and residues with polar side chains, such as asparagine, serine and glutamine. The distribution of charged residues is asymmetric, with more positively charged amino acids associated with the V_H domain than with the V_L domain.

The structure of the hypervariable loops of Jel 72 is illustrated in Figure III-14. Five of the six loops (L1, L2, L3, H1 and H2) belong to canonical structure classes that have been observed for the hypervariable regions of other Fab fragment structures (see Chapter IV). The CDR H3 hypervariable loop of Jel 72 is 13 residues long and rich in basic amino acids, with the amino acid sequence (see appendix) for residues H94 to H102 containing two consecutive glycine residues and three arginines in a span of four residues. This loop folds back and interacts with residues in the L1 and H1 CDRs near the center of the combining region (Figure III-14b). The main chain conformations for the residues at the tip of the H3 loop, Ser H97 to Tyr 100b, is $\beta\alpha_R\alpha_R\alpha_R\alpha_L\alpha_R$. Several of these residues in CDR H3 are involved in maintaining the structural integrity of the loop, including the arginine at position H100. The side chain guanidinium group of this arginine is within 3.5 Å of the side chain of aspartic acid L27d in CDR L1. The side chains of lysine H94 and aspartic acid H101 form a salt link (distance < 3.0 Å apart) at the base of the CDR H3 loop, similar to the interaction observed between an arginine and aspartic acid residue in the structure

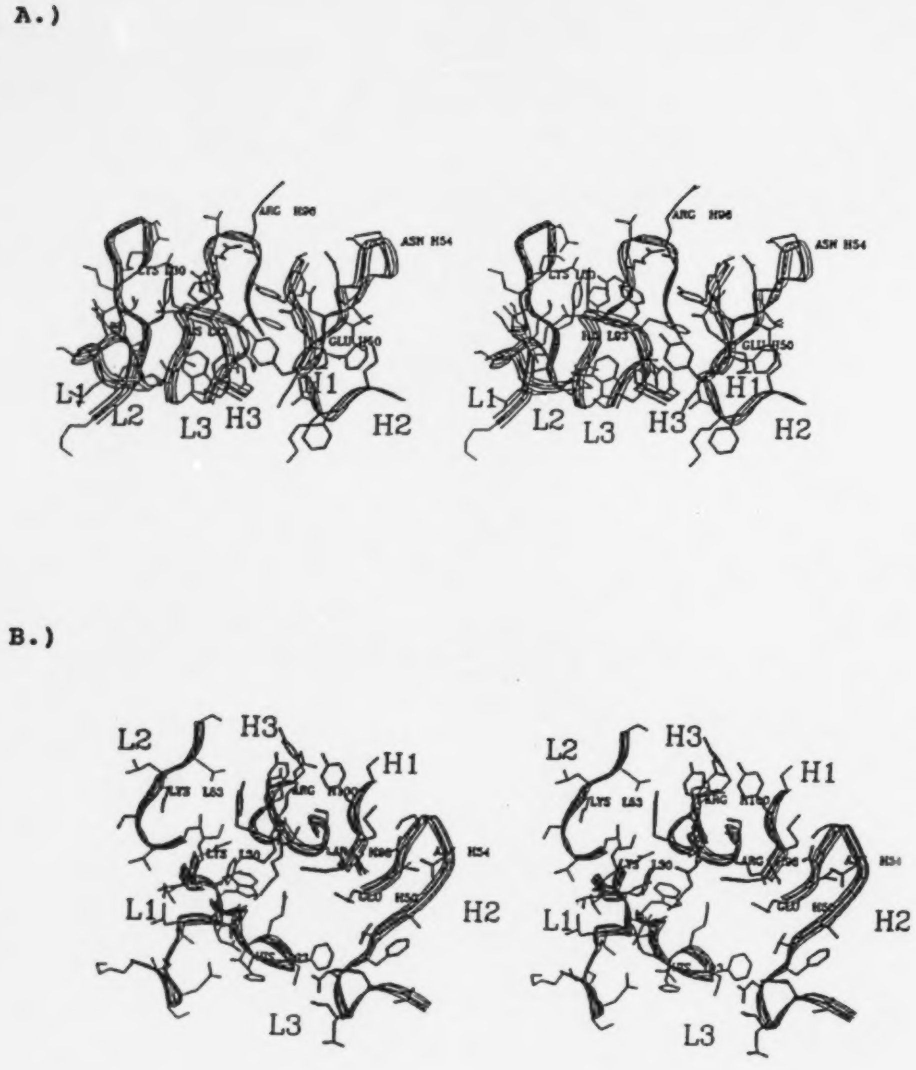

A.)
ARG H96
ASN H54
LYS L30
GLU H50
L1
L2
L3
H3
H1
H2
B.)
L2
H3
H1
LYS L53
ARG
LYS L30
ARG H96
ASN H54
GLU H50
L1
H2
L3

Figure III-14: Stereo views of Jel 72 CDRs. The trace of the main chain is approximated by a ribbon and each CDR is labelled. **A**) view from the L3 side of the combining region, and **B**) view looking directly down onto the combining region.

of Fab McPC603 (Satow et al., 1986). The glycine residue at position H96 has a main chain conformation that would be slightly unfavourable for residues that possess side chains. The Phi/Psi for this residue are near the extended β-sheet region with Phi=-80 and Psi=-150. The presence of the two glycine residues at H95 and H96 may confer increased flexibility to the H3 loop. The temperature factors CDR H3 are relatively high and it seemd likely that this loop will undergo some conformational change upon binding DNA. In the structure of the ssDNA specific Fab BV04-01 (Herron et al., 1991), the CDR H3 loop possessed higher temperature factors in the unliganded Fab than in the complex with $d(pT)_3$, and its conformation changed significantly upon binding the trinucleotide.

The H3 and L1 loops in Jel 72 are the site of strongly positive electrostatic potential (see Chapter V), created in part by the charge on lysine L30, but mostly created by the three arginines in the H3 loop. The occurence of arginines in the CDR H3 loops of autoimmune DNA binding antibodies isolated from mice has been correlated with an increased binding affinity for double stranded DNA by these antibodies (Radic et al., 1989). In this respect, Jel 72 may serve as a good model for other double stranded DNA specific antibodies, because it also possessess a long arginine rich H3 loop. The potential roles these arginines may be playing in recognizing poly(dG)·poly(dC) is addressed in Chapter V.

III.3 Determination and Refinement of Jel 72 (form II)

The crystal of Jel 72 with only one Fab fragment in the asymmetric unit was obtained after the refinement of the first crystal form was nearly complete. This crystal was grown from a drop containing the Fab and the 12 base pair synthetic DNA oligonucleotide d(CG_9CA)·d(GC_9GT). We were thus anxious to determine if this was indeed a crystal of the Fab-DNA complex. Upon determining the space group to be orthorhombic $P2_12_12_1$, with cell dimensions, a=90.56, b=140.71, and c=37.44 Å, a quick calculation of V_m (Matthews, 1968) gave a value of 2.39 $Å^3$/dalton that is identical to the value of 2.40 $Å^3$/dalton for Jel 72 in crystal form I. Both of these values are in the range observed for other protein crystals.

Data Collection: A 2.9 Å data set was collected on a multiwire area detector from San Diego Multiwire Systems, with a Rigaku rotating anode X-ray source. The data collection statistics for this crystal form are presented in Table III-9. The data are weak and extend only to moderate resolution, with a high merging R value between symmetry related reflections. An upper estimate of the overall temperature factor, <B>, obtained from a Wilson plot, is 27 $Å^2$. Nevertheless, a determination of the crystal structure by molecular replacement was attempted.

Table III-9: Data Collection Statistics for Jel 72(II)

Measurements		20,011	
Unique Reflections		8,856	
Resolution (Å)		2.9	
Reflections	d_{min} Å	I/σI	R_{merge}[1]
1,885	4.96	14.5	9.9
1,776	3.94	7.4	12.2
1,765	3.44	3.9	15.7
1,720	3.12	2.2	19.6
1,710	2.90	1.6	20.9
Total		6.1	12.5
Completeness (%)	86		
Completeness (%) $I/\sigma > 2$	49		

1 - R_{merge}, defined as in Table III-1

Molecular Replacement:

We were fortunate in our application of the molecular replacement technique to this crystal form, because we were in possession of excellent models for the variable and constant regions of Jel 72 from the first crystal form. Even with this advantage the interpretation of the rotation function results was not straightforward. Rotation searches were calculated using the Euler angles (θ_1, θ_2, θ_3) as defined by Rossman and Blow (1962) and the procedures as implemented in version 2.1 of X-PLOR (Brünger et al., 1989). The results of these calculations are presented in Table III-10. The rotation function for the constant domain pair gave a clear peak, while the rotation searches with the variable domain pair were more difficult to interpret. We, thus, first positioned the constant domain pair via the translation search. The translation search employed was a strict correlation search. It positions the oriented molecule at different points within the unit cell and computes a correlation coefficient between the observed and calculated data. The oriented constant domain gave a clear result, from which we were able to deduce the correct orientation for the variable domain pair. This orientation was the fifth highest peak in the rotation searches. The poor discrimination between the correct peak and the background noise in the rotation searches with the variable domain pair may be a function of the quality of the data.

Table III-10: Rotation and Translation Search Results for Jel 72(II).

	V Domains	C Domains
Rotation Function		
Resolution (Å)	4 - 10	4 - 10
P.I. Radius[1] (Å)	5 - 24	5 - 24
θ_1	212.6°	198.2°
θ_2	37.4°	39.7°
θ_3	70.7°	77.6°
Peak Height[2]	0.98	1.02
Translation Search		
Resolution (Å)	4 - 15	4 - 15
x	11.6	56.9
y	1.0	71.4
z	20.3	18.1
Peak Height	1.09	1.31

1 - P.I. Radius, Patterson Integration Radius

2 - Peak heights are given as the ratio of the correct peak to the first spurious peak

Structure Refinement:

Refinement of this structure was carried out using X-PLOR as described for crystal form I of Jel 72. First the molecule was divided into the four domains V_H, V_L, C_H and C_L and their orientations and positions optimized by rigid body refinement. Several rounds of molecular dynamics refinement and energy minimization were performed as described for crystal form I of Jel 72. Assignment of individual temperature factors to the atoms resulted in a decrease of the crystallographic residual to 0.203 for 6,258 reflections with $I/\sigma I > 2$ in the resolution range 6.0 to 2.9 Å (Table III-11).

This structure was then inspected with molecular graphics to determine if it contained any bound oligonucleotide. The electron density for the Fab was clear and fit the model well. No electron density for the DNA could be discerned in the combining region, and the packing of the Fabs did not allow for any DNA to fit between symmetry related molecules. The refinement was terminated at this stage because of the low resolution and poor quality of the data. No further work was performed to improve this structure and it is not intended that this structure be published. This structure is of interest, however, because its global packing arrangement is similar to that observed for Jel 318 (Section III.4)

Table III-11: Refinement Statistics for Jel 72(II).

Refinement

	Heavy	Light	Total
Residues	222	219	441
number non-hydrogen atoms	1,679	1,706	3,366

Resolution Range (Å)	Reflections	Shell R-value	Accum. R-value
4.78 - 6.00	852	0.2400	0.2400
4.18 - 4.78	932	0.1970	0.2167
3.80 - 4.18	871	0.2276	0.2201
3.53 - 3.80	846	0.2307	0.2224
3.32 - 3.53	771	0.2424	0.2254
3.15 - 3.32	703	0.2592	0.2293
3.02 - 3.15	658	0.2570	0.2318
2.90 - 3.02	625	0.2615	0.2341
Total Reflections $I > 2\sigma I$	6,258		
Completeness (%)	48		

Geometry

	RMS	Violations
Bond Lengths	0.019	30 > 0.06
Bond Angles	4.425	76 > 12
Dihedral Angles	29.524	0 > 90
Deviations from Planarity		
Improper Angles	1.854	0 > 20
Peptide Bond	2.651	0 > 15

Crystal Packing:

The packing of the Fabs in this crystal form conforms more closely to the crystal packing interactions commonly observed for Fab fragments. The Fabs pack in a "head-to-tail" fashion with the variable regions contacting the COOH termini of a symmetry related constant domain pair. The smallest unit cell dimension is only slightly larger than the thinnest dimension of an Fab fragment, constraining the Fabs to lie roughly in a plane perpendicular to this cell dimension. The elbow angle of the Fab fragment in this crystal form is 138°, which is approximately midway between the elbow angles observed for Jel 72 in the first crystal form.

Comparison with Crystal Form I:

The pseudo-twofold rotation angles relating the variable and constant domains are similar to those observed for the first crystal form of Jel 72. The variable domains are related by a rotation of 173.6° and the constant domains by a pseudo-twofold angle of 166.1°. The RMS deviations from alignment for the Cα atoms are: C_H, 0.91 and 0.69 Å; C_L, 0.65 and 0.66 Å; V_L, 0.68 and 0.65 Å; V_H, 0.89 and 0.82 Å, versus the appropriate domains of Fab-1 and Fab-2, respectively, of crystal form I. The quality of the data (Table III-9) and of the model (Table III-11) is insufficient to present a more detailed analysis of the structure.

III.4 Determination and Refinement of Jel 318

When the space group of the Jel 318 crystals was determined it was noted that the unit cell dimensions were nearly identical to those of the previously determined structure of the Fab fragment Kol (Matsushima et al., 1977). Attempts at solving the structure of Jel 318 on the assumption that it was positioned within the crystal in the same manner as Kol were unsuccessful. Once it was realized that the positioning of the Fab fragment relative to the crystallographic origin was in fact quite different in the two crystals, the molecular replacement solution of Jel 318 became self-evident.

The determination of the rotation and translation parameters for the molecular replacement solution of Jel 318 was achieved by Dr. Miroslaw Cygler. Rotation functions were calculated as described for crystal form I of Jel 72, except that all of the search models were superimposed onto the appropriate domains of Fab Kol. The latter Fab fragment was first positioned with its long axis along **x** and the elbow axis along **y**. Thus, all of the search models possessed an elbow bend angle equivalent to Fab Kol. The initial molecular replacement model consisted of the variable domain pair of the J539 Fab fragment (Suh et al., 1986) and the constant domain pair from Hed10 (Cygler et al., 1987). The translation vector for each domain pair was calculated by a correlation function search (BRUTE; Fujinaga and Read, 1987) and confirmed by calculations of the **S** function (Cygler and Desrochers, 1989). The **S** function performs the translation search in Fourier space rather than Patterson space. The space group is expanded to P1 and an electron density map calculated with phases from a model that is correctly oriented but arbitrarily positioned within the unit cell. A search molecule is then created by applying an appropriate symmetry transformation and moving the model within the electron density map. At each position the score function (**S**) is calculated:

$$\mathbf{S}(p,q,r) = \sum_{i=1}^{N} \mathbf{p}(x_i+p,\ y_i+q,\ z_i+r)$$

The summation is calculated over all **N** atoms, (x_i, y_i, z_i) is the grid point of the electron density map that is closest to the **i**th atom, and (**p**, **q**, **r**) is the translation vector applied to the search molecule, and **p** is the electron density. A maximum in the **S** function indicates the position of the symmetry related molecule, which defines the location of the symmetry element and, thus, the position of the crystallographic origin. The results of the translation searches obtained in Dr. Cygler's laboratory were verified by me using the molecular replacement protocols as implemented in X-PLOR (Table III-13).

Data Collection:

The data were collected from a single crystal using a multiwire area detector from San Diego Multiwire Systems (Xuong et al., 1985). The data were collected and reduced using the UCSD program package (Howard et al., 1985) to produce structure factor amplitudes (F's) from the raw detector images. The data included 97,798 observations of 12,656 unique reflections, R_{merge}= 6.9% (R_{merge}= Σ_{hkl} $|<I>-I_{obs}|/$ Σ_{hkl} $<I>$), representing 92% of the observable data to 2.6 Å resolution with $I > 2\sigma I$. The crystals are sensitive to radiation damage and the quality of the measured intensities decreases rapidly at resolutions beyond 2.8 Å. The quality of the data in the 2.8 - 2.6 Å shell was extremely poor with an unacceptably high merging R value (Table III-12). These data, as well as and additional 310 reflections for which F_{obs}/F_{calc} or F_{calc}/F_{obs} exceeded 2.5, were not included in the refinement. This latter criterion was introduced to eliminate some bad reflections that were not rejected during the merging step. The occurence of such bad errors in area detector data sets can be common and eliminating them from the data set can improve the quality of the refinement (Derewenda et al., 1992). A Wilson plot (Wilson, 1949) of the data yielded an upper estimate of the overall temperature factor, <B>, = 53 $Å^2$. The crystals belong to space group $P2_12_12_1$, with refined unit cell parameters, a = 82.89, b = 139.19, c = 40.96 Å.

Table III-12: Data Collection Statistics for Jel 318

Measurements	97,798
Unique Reflections	12,656
Resolution (Å)	2.8

Reflections	d_{min} Å	$I/\sigma I$	R_{merge}[1]
3,015	4.45	61.4	7.07
3,010	3.53	40.0	8.84
2,981	3.08	13.7	10.94
2,931	2.80	6.2	17.81
2,880	2.60	3.5	26.50
Total		25.3	8.57

Completeness (%)	98
Completeness (%) $I/\sigma > 2$	92

---

1 - R_{merge}, defined as in Table III-1.

Molecular Replacement:

The structure of Jel 318 was solved by the molecular replacement method (Rossman, 1990). Following our experience with the molecular replacement solutions of the Hed10 Fab fragment (Cygler et al., 1987) and Fab Jel 72 , we have used either the V_L-V_H or C_L-C_H domain pairs of Fab fragments as separate search models. We have also noted before that, despite a great structural similarity, there is enough variation between individual domain pairs of different Fab fragments to warrant trials with all available models.

Prior to performing the cross rotation searches, the variable and constant domain pairs of the search models were superimposed on the appropriate domain pairs of Fab Kol (Matsushima et al., 1977). The latter molecule was first oriented with its long axis along X and the elbow bend axis approximately along Y. Rotation searches were calculated with a variety of resolution ranges and Patterson integration radii, as described for Jel 72 (I). Of the Fab domain pairs tried, the clearest rotation function results were obtained using the V_L-V_H domain pair of Fab J539 (Suh et al., 1986), and the C_L-C_H domain pair of Fab Hed10 (Cygler et al., 1987). The translation vectors for these domain pairs were determined by a correlation coefficient search (BRUTE,Fujinaga and Read, 1987), and confirmed by calculations of the **S** function (Cygler and Desrochers, 1989). Because of continued difficulties in the refinement of the C_L-C_H domain pair, the rotation angles and translation vectors for both domain pairs were verified using the molecular replacement protocols as implemented in the X-PLOR program package (Brünger et al., 1987). These results are listed in Table III-13. The correctness of the solution was supported by the self-consistency of the two sets of numbers, and upon the transformation of both domain pairs, their association resembled that of other known Fab fragment structures.

The molecule was then divided into the four individual domains, V_H, V_L, C_H and C_L, and their orientations and positions refined by an interactive local six-dimensional search (Fujinaga and Read, 1987). The correlation coefficient between $|F_{obs}|^2$ and $|F_{calc}|^2$ for a model consisting of 3159 atoms and data in the 4 to 8 Å resolution shell was, at this stage, 0.58 and the R value for data up to 3 Å resolution was 0.43.

Table III-13: Molecular Replacement Results for Jel 318

	C_H-C_L Domain Pair	V_H-V_L Domain Pair
Model Used	Jel 72	Jel 72
Atoms	1,588	1,778
Rotation Search		
Resolution (Å)	4.0-10.0	4.0-10.0
Integration radius (Å)	5.0-24.0	5.0-24.0
Peak[1]	4.85	5.07
θ_1	182.2	206.8
θ_2	50.2	51.6
θ_3	86.0	80.6
Translation Search		
Resolution (Å)	4.0-15.0	4.0-15.0
Peak	8.58	8.44
x Å	50.4	49.8
y Å	102.4	102.4
z Å	32.5	31.3

1, Peak heights are given in standard deviations above background.

Structure Refinement

Prior to the determination of the amino acid sequences of the V_L and V_H domains, preliminary refinement of the structure was carried out using PROLSQ (Hendrickson and Konnert, 1981) resulting in a decrease in the R value to 0.32 for data within the resolution range 6.0 to 2.8 Å. At this point the complete amino acid sequences for the variable domains were determined, and the structure of the antibody combining site was modelled (Chpater IV). These model variable domains, and the constant domain pair from Fab Jel 72 (form I), were used as the starting points for molecular dynamics refinement using the X-PLOR system of programs. Rigid body refinement of these four individual domains converged, indicating that they were correctly positioned. Calculation of the diffraction energy term, WA = 216,820, was followed by 160 cycles of Powell conjugate gradient energy minimization, and then molecular dynamics refinement. The simulated annealing was carried out by heating to 3000 K and slowcooling to 100 K with a step size of 25 K.

At this point the entire structure was inspected with omit maps (Bhat, 1988) and 2 F_{obs} - F_{calc} electron density maps and manually readjusted with interactive molecular graphics using the program INSIGHT (Biosym Technologies, Inc.) on an Iris Silicon Graphics Workstation. The electron density was clear for most of the variable domain pairs but weak and discontinuous for long stretches of surface loops in the constant domain pairs.

For the map calculations, contiguous stretches of approximately 10 - 15 amino acids were omitted from the structure, and energy minimization performed, prior to the calculation of structure factors. The structure was also inspected with $2F_{obs} - F_{calc}$ electron density maps calculated using map coefficients to minimize model bias (Read, 1986). The residues comprising the antibody complementarity determining regions (CDRs) were each, in turn, omitted from the structure for a round of positional refinement and manual refitting. Iterative cycles of molecular dynamics refinement and manual rebuilding resulted in a further dimunition of the R value to 0.27.

The C_H domain was then rebuilt to correspond to the correct heavy chain isotype (IgG2b) and the structure subjected to another cycle of molecular dynamics refinement. Inspection of Ramachandran plots (Ramachandran et al., 1963) indicated problem areas of the structure that were manually rebuilt if necessary. These problem areas were mostly associated with the solvent exposed loop regions of the constant domains, which had higher temperature factors and associated weak electron density. Further cycles of positional refinement and refinement of individual isotropic temperature factors decreased the R value to 0.21 for data in the resolution range 6.0 to 2.8 Å. Since the quality of the electron density was insufficient in the poorly determined regions of the molecule and there was no further improvement in the R value, the refinement was halted at this stage.

This final refined structure consists of 3,301 non-hydrogen atoms and has an R value of 0.207 for 9,342 reflections with $I > 2.5\sigma I$, representing 86% of the observable reflections for data in the resolution shell 6.0 to 2.8 Å. The refinement statistics and the root mean square deviations from ideality of the geometric parameters is presented in Table III-14. A Ramachandran plot of the main chain torsion angles for the Heavy and Light chains of this structure is presented in Figure III-15. The temperature factor distribution, averaged over all atoms for the residues of both chains, is shown in Figure III-16. Representative regions of good and poor electron density are displayed in Figure III-17. The error in the coordinates, as estimated by the method of Luzzati (Luzzati, 1952), is between 0.30 and 0.40 Å (Figure III-18). Although there is some electron density in the final electron density maps that could be interpreted as water molecules, no ordered solvent has been included in the final model. Further refinement of this structure would require the collection of better quality, higher resolution data.

Table III-14: Refinement Statistics for Jel 318

Refinement	Heavy	Light	Total
Residues	219	213	432
number of non-hydrogen atoms	1,662	1,639	3,301

Resolution Range (Å)	Reflections	Shell R-value	Accum. R-value
4.68 - 6.00	1,314	0.1778	0.1778
4.07 - 4.68	1,298	0.1570	0.1674
3.69 - 4.07	1,284	0.2042	0.1775
3.42 - 3.69	1,203	0.2023	0.1823
3.21 - 3.42	1,194	0.2219	0.1877
3.05 - 3.21	1,088	0.2503	0.1938
2.91 - 3.05	998	0.2695	0.1992
2.80 - 2.91	963	0.2889	0.2043
Total Reflections $I > 2.5\sigma I$	9,342		
Completeness (%)	86		

Geometry	RMS	Violations
Bond Lengths	0.014	9 > 0.060
Bond Angles	3.997	50 > 12
Dihedral Angles	28.134	0 > 90
Deviations from Planarity		
Improper Angles	1.89	0 > 20
Peptide Bond	2.80	0 > 15

A.

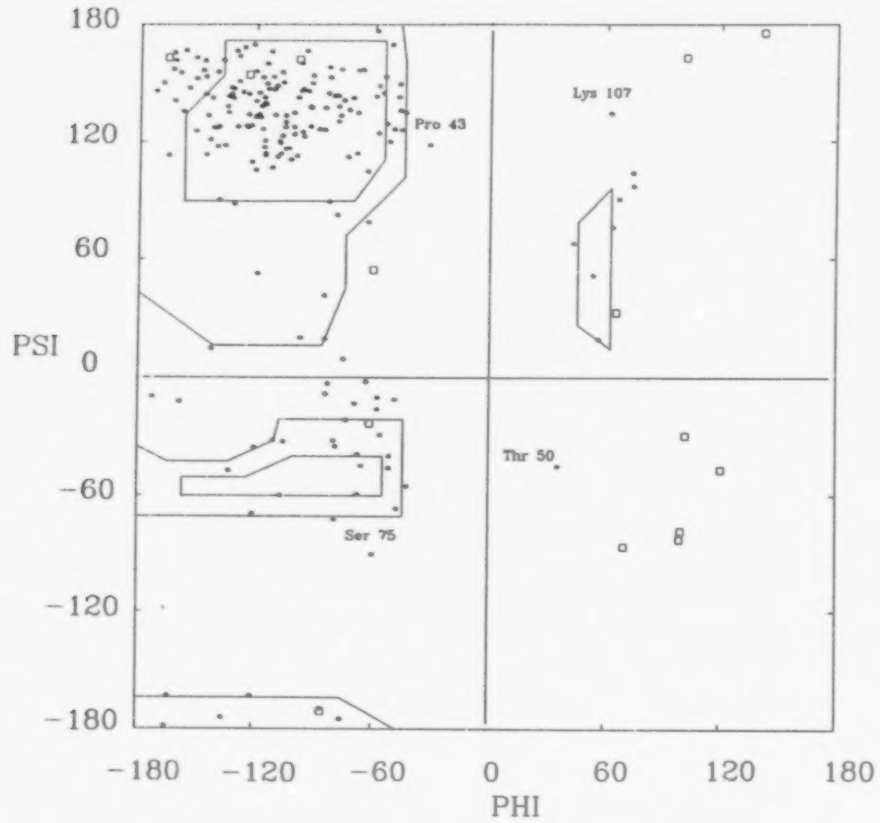


Figure III-15: Ramachandran Plot of the main chain torsion angles (Phi/Psi) for the residues of Jel 318. The positions of the glycine residues are indicated by squares, all other residues by circles. The residues that lie outside of the allowed regions are labelled according to their Kabat residue number (Kabat et al., 1987). **A)** Light Chain of Jel 318.

B.

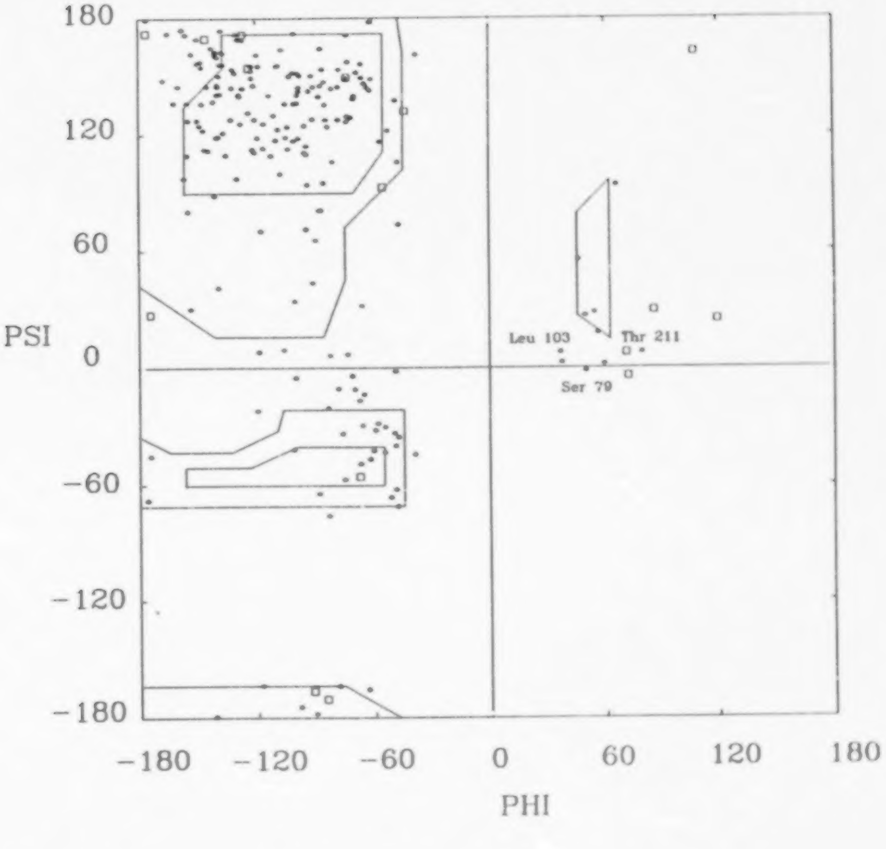


Figure III-15: Ramachandran Plot of the main chain torsion angles (Phi/Psi) for the residues of Jel 318. The positions of the glycine residues are indicated by squares, all other residues by circles. The residues that lie outside of the allowed regions are labelled according to their Kabat residue number (Kabat et al., 1987). **B)** Heavy Chain of Jel 318.

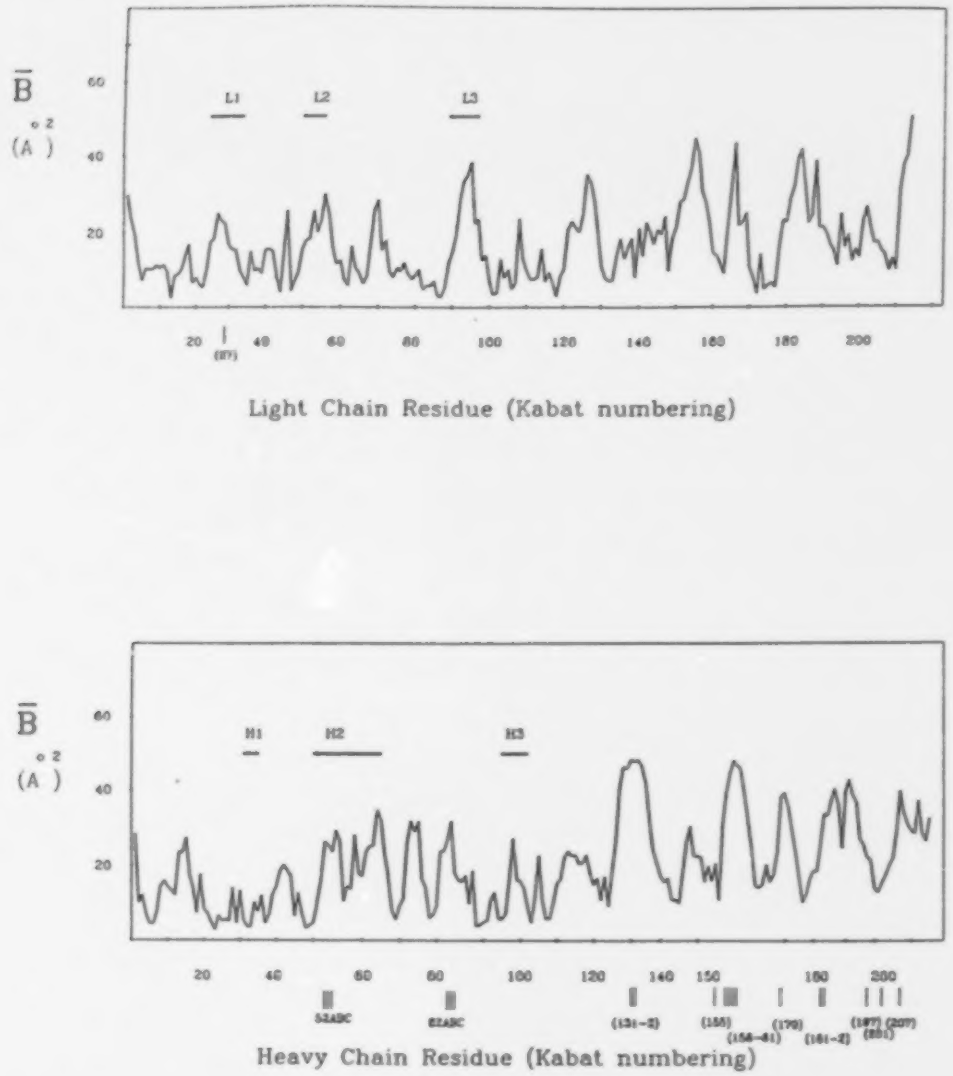


Figure III-16: Plots of mean temperature factor by residue for the Light (top) and Heavy (bottom) chains of Jel 318. The numbering of Kabat et al. (1987) is used for both chains with the positions of insertions marked A, B, etc and deletions shown in parantheses. The positions of the antibody CDRs are indicated by the solid bars.

Figure III-17: Representative $2F_{obs}-F_{calc}$ electron density maps contoured at 1σ showing region of good density for residues H52 - H53 of CDR H2 (top) and poor electron density for residues H129 - H134 of the C_H domain near the interchain disulphide (bottom).

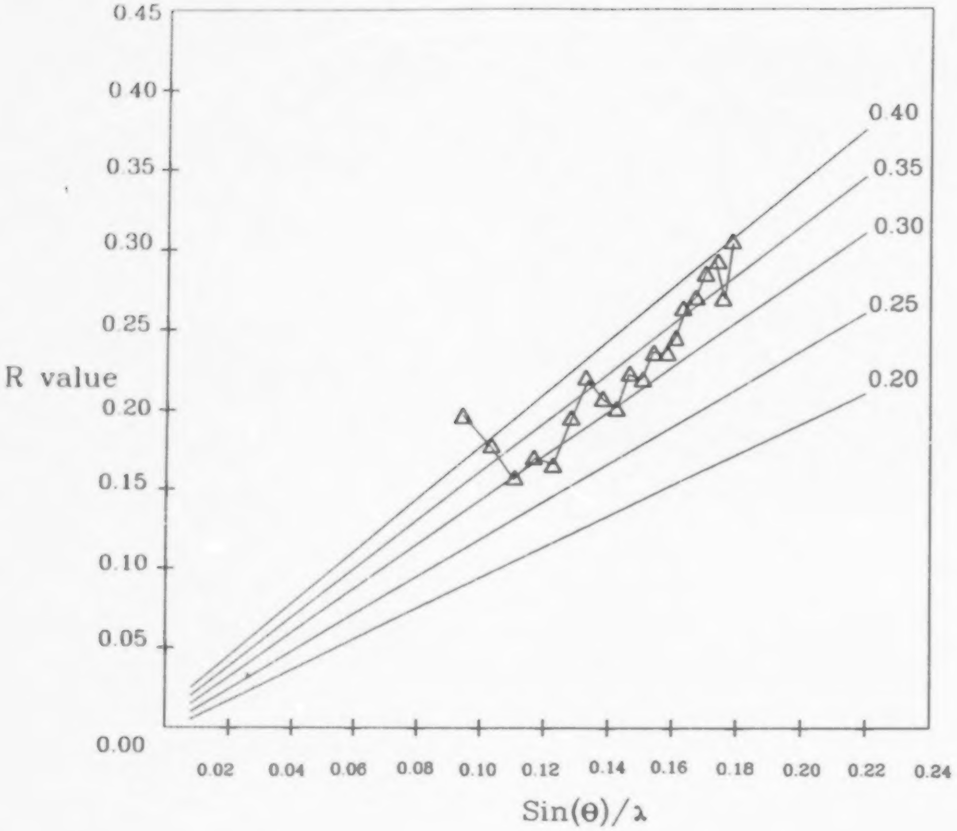


Figure III-18: Plot of R value (triangles) by shell on reflections for which $I > 2.5\sigma(I)$, versus sin(θ)/lambda for Jel 318 , along with the theoretical curves (Luzatti, 1952) for the R value if all discrepancy is attributable to the coordinate error.

Crystal Packing:

The crystal packing of Jel 318 is similar to that observed in the second crystal form of Jel 72. In both crystals the variable regions of the Fab fragments contact the constant regions of symmetry related molecules. The crystal packing interactions of Jel 318, however, are predominantly between residues of the V_H domain and residues near the carboxy terminus of a C_L domain from a symmetry mate. The crystal contacts for Jel 318 are listed in Table III-15. The table lists those residues that come within a 4.0 Å radius of symmetry related residues. Portions of all three heavy chain hypervariable regions are involved in crystal contacts with a symmetry C_L domain. Unlike the carboxy terminal residues of the heavy chain in the form I crystals of Jel 72, the carboxy terminus of the C_H domain of Jel 318 is not involved in any crystal contacts. This may explain why crystals of Jel 318 could be grown from different isoelectric species of Fab fragments and why the crystallization was relatively insensitive to the isoelectric purity, or lack thereof, of the Fab preparations (Chapter II).

The repeating lattice of the Jel 318 crystals is in the shape of a long shallow box, with the unit cell dimensions of Jel 318, a=82.0, b=139.2, c=41.3 Å, being similar to those of the second crystal form of Jel 72 and nearly identical to the crystals of Fab Kol (Matsushima et al., 1977). Another Fab fragment, J539 (Suh et al., 1986) crystallizes in a unit cell

Table III-15: Crystal Contacts for Jel 318

L :	**H :**	Symmetry Mate
14-18,22,107	3,5,7,13, 15,23,25,109	4
H :	**L :**	
31-32,52-53	191,209,211,213-214	3
96-97	191,212-213	3
1	109,200	4
24-27	109-111	4
78	170-171	4
80	109-110	4
52,53,58	129-130	3

Residues within the CDRs are listed in bold type.

Symmetry Mate:
1 - (x, y, z)
2 - (1/2-x, -y, 1/2+z)
3 - (1/2+x, 1/2-y, -z)
4 - (-x, 1/2+y, 1/2-z)

with similar cell dimensions. Table III-16 lists the unit cells of each of these four crystals as well as the elbow angle relating the variable and constant domain pairs within each structure.

Despite the fact the individual cell dimensions can vary significantly and the variation in the elbow angle relating the variable and constant regions is 33°, all four Fabs pack within the crystals in a similar manner. The primary crystal contacts are made between residues of the variable domains near the antibody combining region with residues near the COOH termini of symmetry related constant regions. This type of packing is commonly observed in crystals of Fab fragments (Cygler et al., 1987). The packing differs, however, in the relationship between the relative position of the molecule within the unit cell and the space group symmetry elements. All four crystals belong to the space group $P2_12_12_1$, which is characterized by three non-intersecting screw axes parallel to the crystallographic axes. The positioning of each Fab within its respective unit cell is illustrated in Figure III-19. As is apparent in the figure the packing modes can be divided into two types. Either the crystallographic **a** axis passes between the variable and constant regions of a single molecule, or it passes between two symmetry related molecules. Thus, the constant region of Jel 318 has roughly the same translation vector as the variable region of Kol. Variation in the elbow angle of the Fab is accommodated within the

Table III-16: Unit Cell Dimensions of Fab Fragment Crystals Similar to Jel 318

	a	*b*	*c*	*elbow angle*
Jel 318	82.0	139.2	41.3	158°
Kol	78.3	138.9	40.0	170°
Jel 72(II)	90.6	140.7	37.4	137°
J539	74.2	130.8	54.1	145°

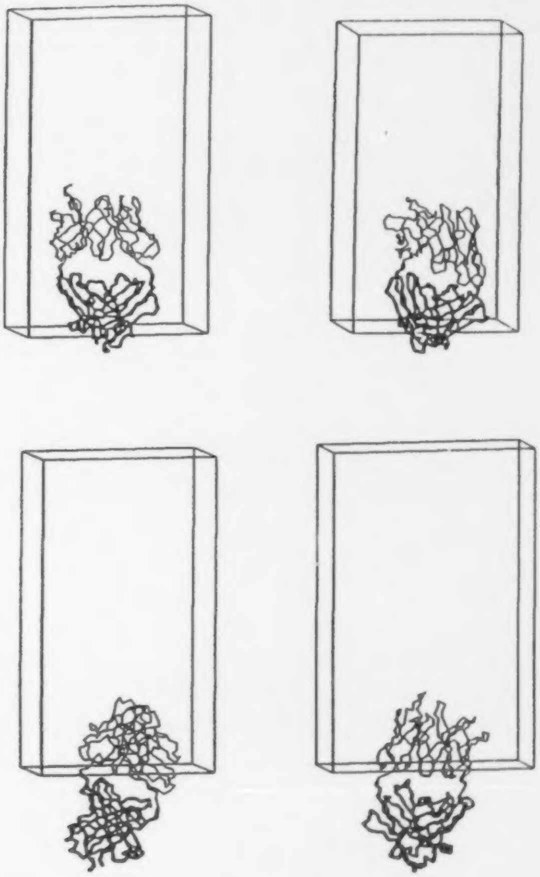


Figure III-19: Differences in crystal packing interactions among similar crystals. For each Fab the constant domain pair is drawn in thicker lines than the variable domain pair. The figure depicts each Fab packing within its own unit cell: Jel 318 (upper left), Kol (lower left), J539 (upper right), and Jel 72 (form II) (lower right).

crystal lattice by altering the width or depth of the unit cell. The compact elbow angles of J539 and Jel 72 are accommodated by a slight twist of the Fab about its long axis relative to the crystal axes. In this way, the global packing arrangement of Jel 318 more closely resembles J539, while the packing of Jel 72 is more similar to Kol, despite the fact that their unit cells are not isomorphous and their is a large variation in their elbow angles. The fact that the unit cell dimensions of the crystals of Jel 318 are nearly identical with those of Fab Kol complicated the interpretation of the translation function results. This experience illustrates that when attempting the molecular replacement method one should not allow preconceived notions based on such things as apparently isomorphous crystals preclude the possiblilty that the crystal packing may be quite different in the molecule under investigation. We have observed a similar situation in orthorhombic crystals of a lysozyme-peptide chimaeric protein, where the lysozyme also packs differently relative to the origin despite unit cell dimensions that are very similar to the orthorhombic crystal form of Hen egg white lysozyme (Artymiuk et al., 1982; H. Patel and W. F. Anderson, unpublished results).

Structure Quality

The structure of Fab Jel 318 contains the immunoglobulin fold that has been observed in all other Fab fragment crystal structures. The elbow angle relating the variable domain pair to the constant domain pair is 154° and is approximately in the middle of the range of elbow bends observed for Fab fragments. The rotation and translation operations that relate the V_L to the V_H domain (171°, 0.25 Å) and the C_L to the C_H domain (168°, 1.8 Å), are also similar to other known Fab structures (Cygler and Anderson, 1988).

There is good, general agreement between the model and the electron density in $2F_{obs} - F_{calc}$ electron density maps contoured at the 1σ level. As is commonly the case in Fab fragment structures, this correspondance is generally clearer for the Light chain than the Heavy chain, and for the variable domains compared to the constant domains, a fact that is reflected in the temperature factor distributions for these domains (Figure III-16). The electron density is clear for the main chain and side chains of most of the residues within the variable region CDRs. As can be expected with data at this resolution, the fine details of some of the side chain torsion angles are not as well determined, particularly in those regions of the structure that are characterized by large temperature factors. The variations of the residue averaged temperature factors correlates closely with the alternating structure of β-strands and connecting loops along the polypeptide chain. There is weak electron density, and consequently higher temperature factors, for the β-strand connecting loops of the constant domains (residues H160-H165, H195-H200, and L150-L155, L180-L185). Despite the fact that there are crystal packing interactions in the vicinity of the interchain disulphide, those residues of the H chain loop between positions 132-140 and the extreme carboxy terminus of the L chain (211-214) that are not specifically involved in crystal contacts also possess large temperature factors and poorly defined electron density. Poor quality electron density for the interchain disulphide bond is commonly observed in Fab fragment crystal structures (see eg., Tormo et al., 1992; Tulip et al., 1992), and is usually attributed to the fact that this disulphide is on the surface of the molecule and may exist in multiple conformations.

A few residues possess main chain torsion angles that lie outside the allowed regions of the Ramachandran plot (Figure III-15). These residues are in solvent exposed loops associated with regions of poorly defined electron density, and are for the most part confined to the constant domains, such as C_L ASP 151 and C_H GLU 150 and ASN 162. Of the residues that are part of the antibody CDRs, the outliers on the Ramachandran plot includes residue Leu H103, which lies just outside of the left handed α-helical region. These outlying residues have well-defined density for their main chain and side chain atoms that appears to confirm their conformations.

Similar distortions are not uncommon and have been observed in other Fab fragment structures (see eg., Tulip et al., 1992; Stiepe et al., 1992; and Saul and Poljak, 1992). The other significant CDR residue outlier is Thr L50 which is in a region that was initially considered forbidden near the region Phi=60°, Psi=-30°. A similar conformation for this residue has been observed in a 2.0 Å resolution structure of a recombinant V_L domain from the antibody McPC603 (Stiepe et al., 1992) and calculations indicate that this region of the (Phi/Psi) plot is in fact favoured if there is some flexibility allowed for the atomic bonds (Weiner et al., 1984).

The Antibody Combining Region:

The surface that Jel 318 presents for interaction with triplex is remarkably flat, punctuated by the protrusion of a single hypervariable loop, CDR H2, away from the main body of the molecule (Figure III-20). In this respect, the combining region of Jel 318 is similar to that of the dsDNA specific Fab fragment Jel 72, and is unlike the combining regions of the ssDNA specific Fab fragments Hed10 and BV04-01 (Herron et al., 1991). Both of the ssDNA binding Fabs are characterized by combining regions that form deep clefts created by having long CDR L1 and CDR H2 hypervariable loops. Jel 72, however, forms a much less pronounced cleft since it possesses a shorter CDR H2 loop and an exceedingly long CDRH3 loop that folds back into the center of the combining region.

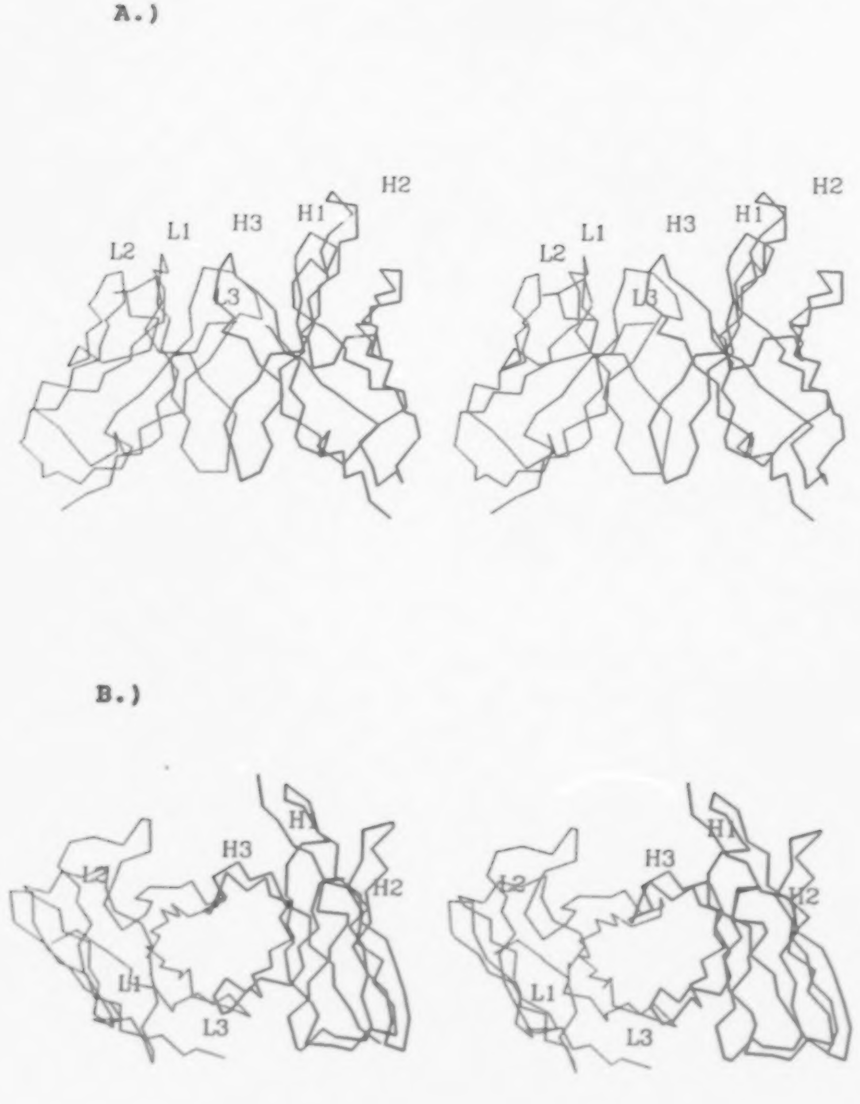
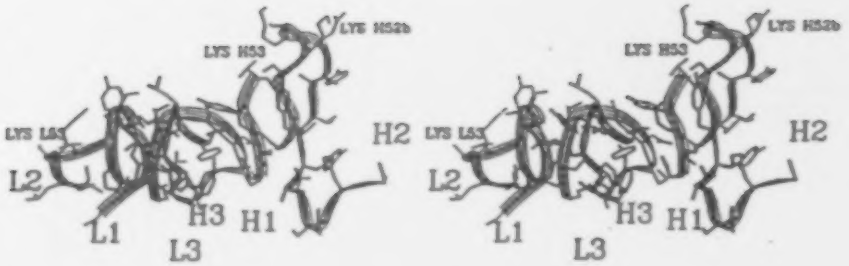

A.)
H2
H1
L1
H3
L2
L3
H2
H1
L1
H3
L2
L3
B.)
H1
H3
L2
H2
L1
L3
H1
H3
L2
H2
L1
L3

Figure III-20: Stereo diagram of the alpha carbon trace of the polypeptide backbone for the $V_H \cdot V_L$ domains of Fab Jel 318. The antibody CDRs are labelled. The V_H domain is shown in bold lines, the V_L domain in thin lines. **A.)** view from the L3 side of the combining region and, **B.)** view looking down into combining region

The relatively flat surface of the Jel 318 combining region is created by having a small CDR L1 loop, shorter by six residues than the same loop observed in either Hed10 or Jel 72, and also a short CDR H3 loop. The resulting combining surface is enriched in amino acids with small polar side chains, such as serine and asparagine residues. Considering the extremely anionic nature of triple stranded DNA it is somewhat surprising that the combining region does not contain more amino acids with basic side chains. The basic amino acids are mostly located in the protruding CDR H2 hypervariable loop, which contains three basic residues, Arg H50, Lys H52b and Lys H53, although there is also a single arginine residue, H98, in CDR H3 and a lysine residue at position L53 in CDR L2. (Figure III-21).

The main chain conformation of five of the six CDRs conforms to that previously observed in Fab fragment structures and to the canonical structure models described by Chothia and Lesk (1987). Thus, CDR L1 belongs to the canonical structure class 4, CDR L3 to canonical class 1, CDR H2 to the canonical class 4, and CDR L2 and CDR H1 to the only canonical structure that has been observed for them. No canonical structures for the CDR H3 region have been derived because the length of this loop and its amino acid sequence is highly variable among Fab fragments of known structure. In Jel 318, CDR H3 occupies a central position in the combining region and interacts with the other hypervariable loops to form the flat

A.)


LYS H52b
LYS H53
LYS L53
H2
L2
H3
H1
L1
L3
LYS H52b
LYS H53
LYS L53
H2
L2
H3
H1
L1
L3

B.)

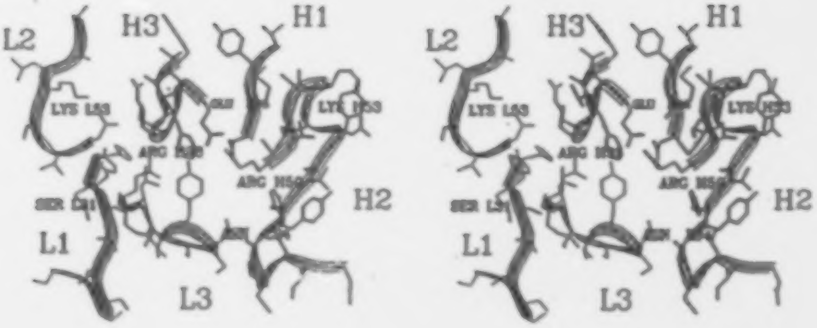

L2
H3
H1
LYS L53
LYS H53
H2
L1
L3
L2
H3
H1
LYS L53
LYS H53
H2
L1
L3

Figure III-21: Stereo views of the antibody CDRs of Fab Jel 318. The trace of the polypeptide backbone is approximated by a ribbon and the six CDRs are labelled. **A.)** View from the L3 side of the combining region and, **B.)** view looking directly down into the combining region.

combining surface. This loop contains seven amino acids with the sequence Glu Leu Leu Arg Ser Phe Ala Tyr from positions H95 to H102 (see Appendix). These residues possess main chain conformations of the form $\beta\beta\alpha_L\alpha_R\beta\alpha_R\alpha_L\beta$.

Several of these residues in CDR H3 interact with residues in the other CDRs. The side chain of glutamic acid at position H95 forms a salt link with the side chain of Arg H50 (distance, 2.8 Å) at the beginning of the protruding CDR H2 loop. The side chain of the arginine residue at position H98 is exposed to solvent, but its Nϵ nitrogen atom is within hydrogen bonding distance (2.7 Å) of the side chain of the first residue in CDR L2, Aspartic acid L50, while the side chain hydroxyl group of serine H99 can also hydrogen bond with the side chain of the framework residue Arg L46. These interactions serve to pack the CDR H3 loop within the center of the combining region, and illustrates the point that at least some of the charged residues in the combining regions of antibodies may play a role in the stabilization of the structure of the CDRS, rather than a functional role in the recognition of antigen. In DNA binding antibodies an assumption is often made that arginine residues within the CDRs necessarily are involved in recognition of the sugar phosphate backbone of the DNA antigen, while they may, in fact, be essential for stabilizing the structure of the hypervariable loops.

III.5 Refinement of Hed10

The structure of the Fab fragment from the ssDNA specific antibody Hed10 was the first one determined in our laboratory. The data were collected and the determination of the rotation and translation parameters for the molecular replacement solution was achieved by Dr. Miroslaw Cygler (Cygler et al., 1987). As with the refinements of both Jel 72 and Jel 318, the refinement of the structure of Hed10 was hampered by the lack of the amino acid sequence for the variable domains. Of the three Fab fragments whose structures we have determined, Hed10 was the first to have its amino acid sequence determined and, consequently, the first to have a model derived for its Fv fragment based on the amino acid sequence of the hypervaraible loops (see Chapter IV).

In order to assess whether the modelling protocol was successful, the structure of Hed10 had to be further refined. In this section this very limited refinement of Hed10 is described and a description of the antibody combining region presented. As with the structure of Jel 72 in crystal form II, this structure will not be published. The structure of Hed10 refined against data to 2.0 Å resolution will be published elsewhere (M. Cygler personal communication) and is not the structure presented here. This structure, refined against data to 2.4 Å resolution, is the structure that is compared with the model structure derived in Chapter IV.

Structure Refinement

Hed10 crystallizes in the monoclinic space group $P2_1$ with refined unit cell parameters, a = 64.2, b = 90.0, c = 42.3 Å, and $\beta = 96.7^\circ$ (Cygler et al., 1987). When the model for the structure of the variable domains of Hed10 was derived, the crystal structure of Hed10 was at a stage of refinement that had been very nearly completed by Dr. Cygler, with an R value for data to 2.4 Å resolution of 0.25. Inspection of this structure with molecular graphics, and with reference to the now known amino acid sequence, revealed that the CDR L1 hypervariable loop had been built with the incorrect sequence and that a portion of the CDR H2 hypervariable loop had been omitted from the structure. Thus, this structure was further refined by molecular dynamics using X-PLOR (Brünger et al., 1987).

Calculation of the diffraction energy term, WA = 293,350, was followed by 160 cycles of Powell conjugate gradient energy minimization and one cycle of molecular dynamics refinement. Omit maps (Bhat, 1988) were calculated, as described previously for Jel 72(I) and Jel 318, for the region of CDR L1 that had been built with the incorrect sequence, residues Ser L26 through Asn L31, and for the region of CDR H2 that had been omitted, residues Ser H52C through Ala H56. With the exception of the side chain density for residue Tyr H55, the density was clear and unambiguous for the trace of the main chain in both hypervariable regions. The electron density for the side chain of Tyr H55 was discontinuous from Cβ onward, however, there was clear density for the edge of the tyrosine ring nearest the OH group and for the tyrosine OH group itself.

The entire structure was then inspected in $2F_{obs} - F_{calc}$ electron density maps contoured at the 1σ level and the model was found to be in good agreement with the observed electron density. Residues with strained backbone conformations, as determined from inspection of Ramachandran plots (ramachandran et al., 1963) were examined in omit maps and $2F_{obs} - F_{calc}$ electron density maps, and adjusted if necessary. Since the data for Hed10 extended to higher resolution, the electron density maps were examined for the presence of bound water molecules. The coordinates for several potential water molecules had been provided by Dr. Cygler, however, many of these waters were rejected on the basis that they were not in electron density, as observed in $F_{obs} - F_{calc}$ electron density maps contoured at the 3σ level, or were not in a position to make hydrogen bonds with suitable groups on the protein. Potential bound waters were also excluded if, after refinement of their individual temperature factors, they possessed temperature factors that exceeded 70 $Å^2$. Iterative cycles of inspecting electron density maps for the presence of bound waters, refineing their positions and temperature factors, and exluding those water molecules that did not meet the above criteria, resulted in coordinates for 76 bound waters.

The refinement statistics for this structure, containing the bound water molecules, is presented in Table III-17. If the bound waters are not included the R value for data between 6.0 and 2.4 Å resolution is 0.220. Ramachandran plots for the heavy and light chains of this structure are presented in Figure III-22. The residue averaged temperature factor distributions for both chains are presented in Figure III-23. The temperature factors for the water molecules vary from a low of 17 $Å^2$ to a high of 68 $Å^2$. The mean coordinate error for this structure, as estimated by the method of Luzatti (Luzzati, 1952) is between 0.25 and 0.35 Å (Figure III-24).

Table III-17: Refinement Statistics for Hed10

Refinement	Heavy	Light	Solvent	Total
Residues	219	219	(76)	438
number of non-hydrogen atoms	1,661	1,685	(76)	3,422
Resolution Range (Å)	Reflections		Shell R-value	Accum. R-value
4.24 - 6.00	2,110		0.1516	0.1516
3.59 - 4.24	2,086		0.1682	0.1590
3.22 - 3.59	1,958		0.1881	0.1664
2.96 - 3.22	1,976		0.2169	0.1746
2.77 - 2.96	1,876		0.2459	0.1823
2.62 - 2.77	1,772		0.2580	0.1883
2.50 - 2.62	1,677		0.2761	0.1939
2.40 - 2.50	1,474		0.2692	0.1978
Total Reflections $I > 2\sigma I$	14,929			
Completeness (%)	85			
Geometry				
	RMS		Violations	
Bond Lengths	0.014		9 > 0.060	
Bond Angles	3.570		36 > 12	
Dihedral Angles	29.073		0 > 90	
Deviations from Planarity				
Improper Angles	1.730		0 > 20	
Peptide Bond	2.726		0 > 20	

A.)

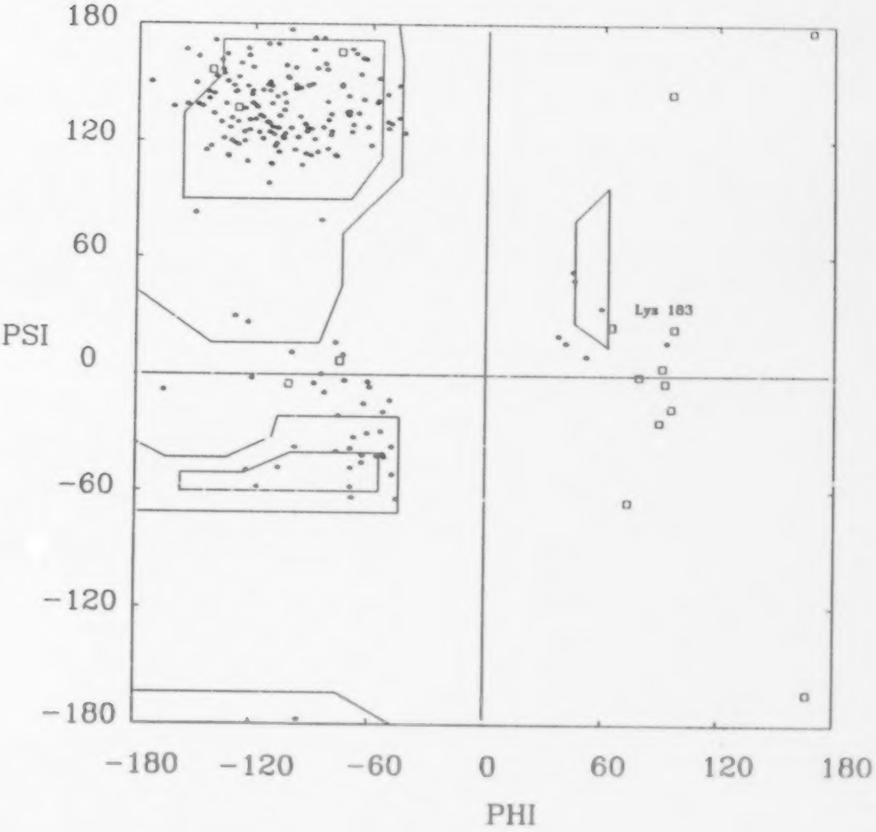


Figure III-22: Ramachandran plot of the main chain torsion angles (Phi/Psi) for the residues of Hed10. The positions of the glycine residues are indicated by squares, all other residues by filled circles. **A.)** Light Chain of Hed10.

B.)

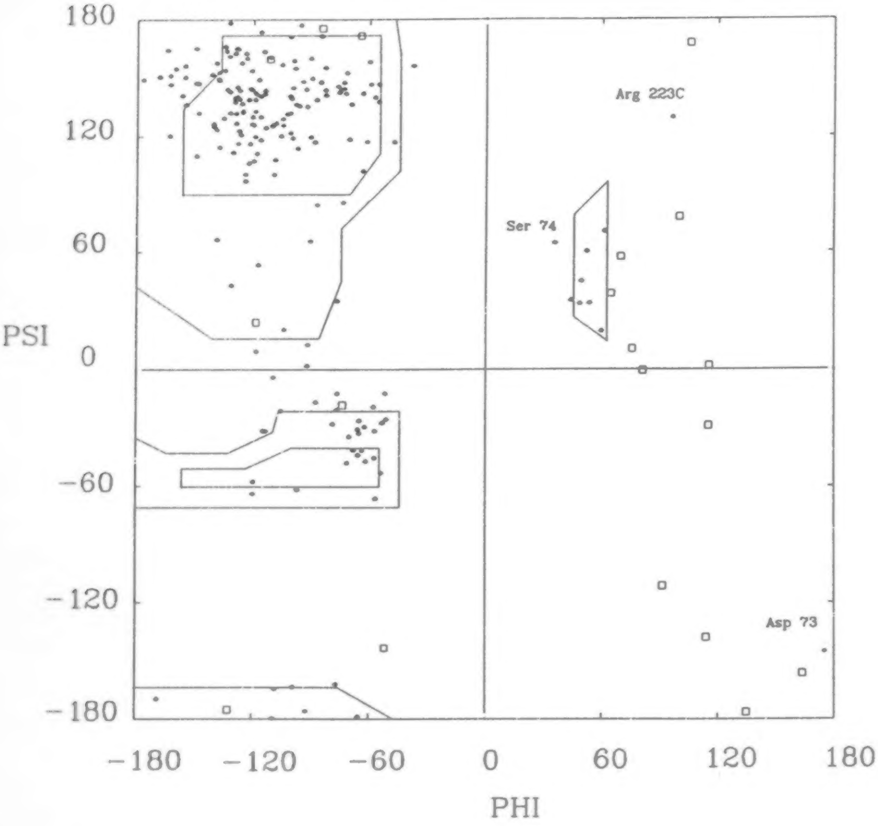


Figure III-22: Ramachandran plot of the main chain torsion angles (Phi/Psi) for the residues of Hed10. The positions of the glycine residues are indicated by squares, all other residues by filled circles. **B.)** Heavy Chain of Hed10.

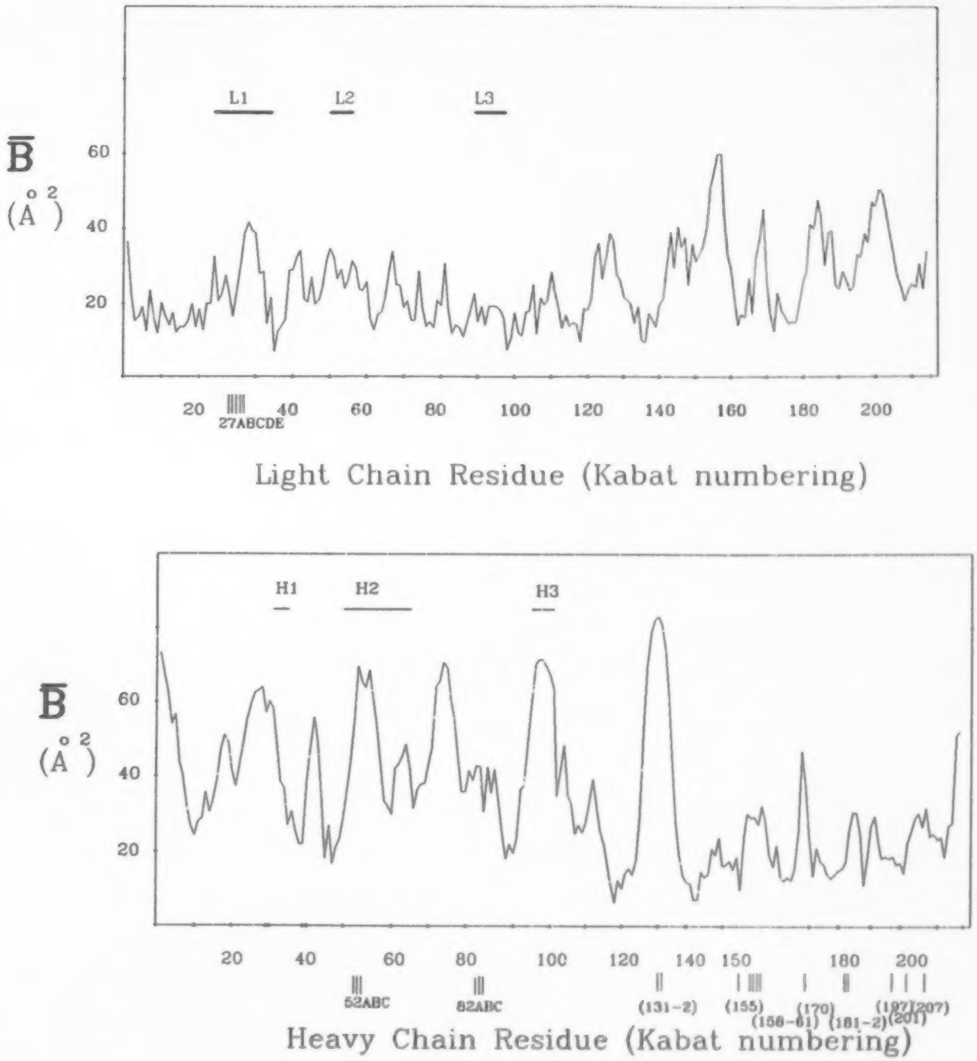


Figure III-23: Plots of mean temperature factor by residue for the Light (top) and Heavy (bottom) chains of Hed10. The numbering of Kabat et al. (1987) is used for both chains with the positions of insertions marked A, B, etc and deletions shown in parentheses. The positions of the antibody CDRs are indicated by the solid bars.

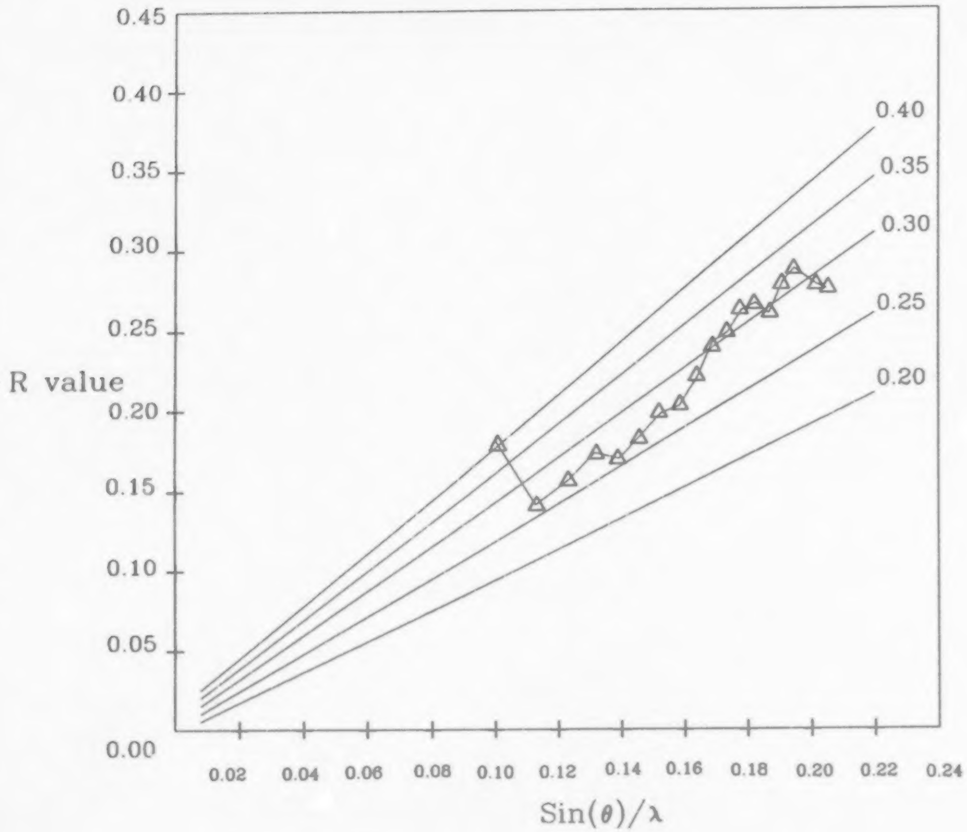


Figure III-24: Plot of R value (triangles) by shell on reflections for which $I > 2.0\sigma(I)$, versus sin(Θ)/lambda for Hed10, along with the theoretical curves (Luzatti, 1952) for the R value if all discrepancy is attributable to the coordinate error.

Structure Description

The structure of the Fab fragment from the murine autoimmune antibody Hed10 contains the immunoglobulin fold that has been observed in both Fab Jel 72 and Fab Jel 318 and in all other Fab fragment structures (Figure III-25). The elbow angle relating the variable domains to the constant domains is 162°, within the range observed in Fab fragment structures. Thus, Hed10 is in a much more extended conformation than either of the three observed Fab fragment structures for Jel 72, and is slightly more extended than Jel 318. The pseudo-twofold rotations that relate the Light chain domains to the Heavy chain domains of Hed10 are also within the range observed for other Fab fragment structures. The rotation relating the V_L domain to the V_H domain is 175.6°, whilt the rotation relating the C_L domain to the C_H domains is 168.8°.

The electron density is clear for the trace of the main chain and the positions of the side chains. The density is poorest for the V_H domain, the loops of which possess higher tmeperature factors than the corresponding loops of the V_L domain. The average temperature factor for the Heavy chain is 30.5 $Å^2$, while for the Light chain it is 21.4 $Å^2$, which also compares favourably with recently reported Fab fragment crystal structures (see eg. Tormo et al., 1992).

As mentioned previously, the antibody combining region of Hed10 forms a deep cleft that is lined with the side chains of aromatic amino acids, particularly the tyrosine residues of the H3 loop. The combining region cleft is created by Hed10 having long loops for CDR L1 and CDR H2, that form the sides of the cleft approximately parallel to the $V_H \cdot V_L$ domain interface (Figure III-26a), and a shorter CDR H3 loop that forms the base of the cleft on the H3 side of the combining region.

Five of the six hypervariable regions of Hed10 possess main chain conformations that place them in the canonical structure classes for immunoglobulin loops described by Chothia and Lesk (1987). The three Light chain hypervariable regions: L1, belonging to canonical class 4 for this loop; L2, to canonical class 1; and L3, to class 1; as well as the first two Heavy chain hypervariable regions: H1, belonging to canonical class 1; and H2, belonging to canonical class 4 for this loop. The H3 hypervariable region forms a β-hairpin loop and contains three tyrosines and two glycine residues. The glycine at position H97 has a main chain conformation that is unfavourable for amino acids that possess side chains, with phi= 90° and psi= -110. The second glycine in the H3 loop at position H101 is in the β conformation. The main chain conformations for all of the residues of the H3 loop from position Tyr H95 to Phe H102, excluding the glycine at H97, is $\beta\beta$-$\beta\beta\beta\beta$.

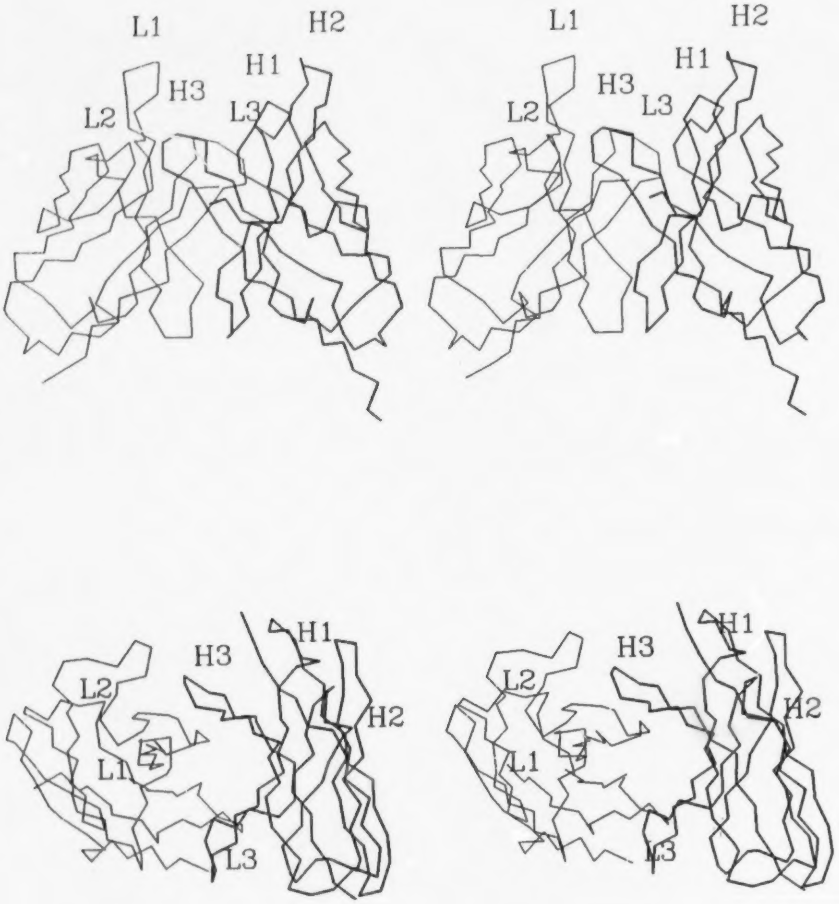

L1
H2
H1
H3
L2
L3
L1
H2
H1
H3
L2
L3
H1
H3
L2
H2
L1
L3
H1
H3
L2
H2
L1
L3

Figure III-25: Stereo view of the structure of the Cα trace of the variable domains of Hed10. The Heavy chain is drawn in bold lines, the Light chain in thin lines. The positions of the antibody CDRs are labelled. View from the L3 side of the combining region (top) and view looking down into the combining region (bottom)

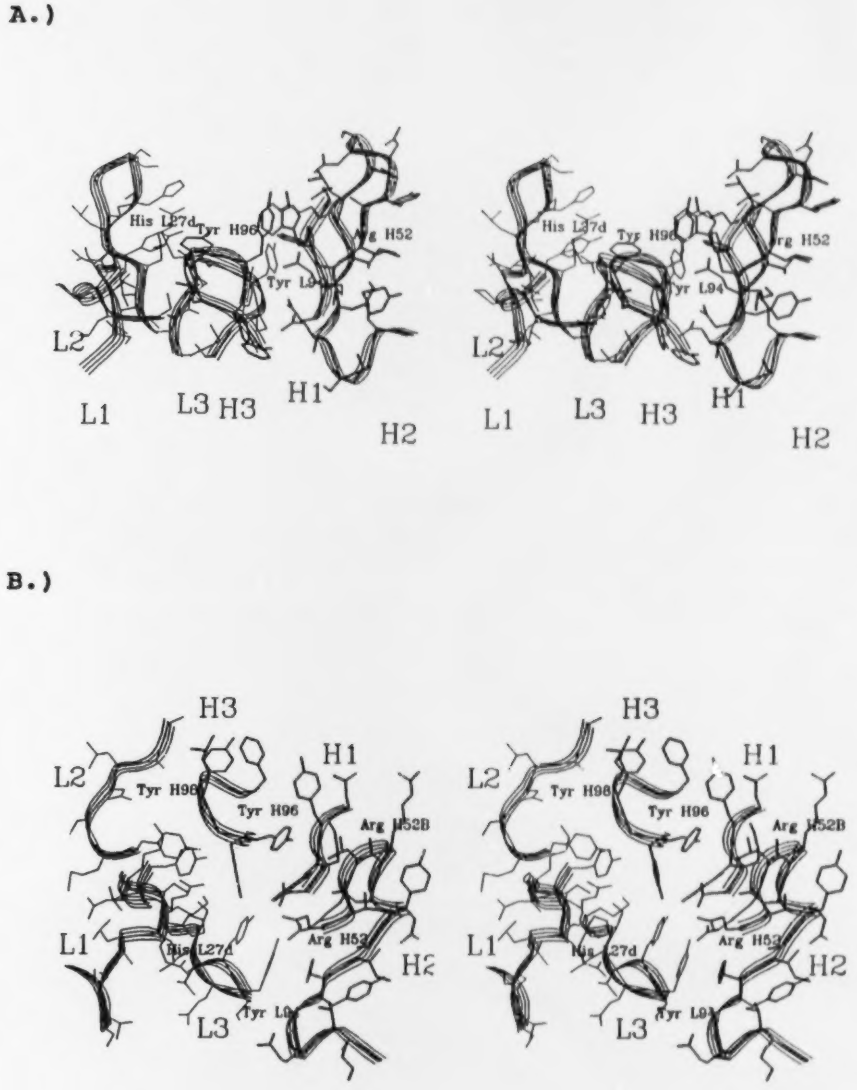

A.)
His L27d
Tyr H96
Arg H52
Tyr L94
L2
L1
L3
H3
H1
H2
His L27d
Tyr H96
Arg H52
Tyr L94
L2
L1
L3
H3
H1
H2
B.)
H3
L2
Tyr H98
Tyr H96
H1
Arg H52B
L1
His L27d
Arg H52
H2
Tyr L94
L3
H3
L2
Tyr H98
Tyr H96
H1
Arg H52B
L1
His L27d
Arg H52
H2
Tyr L94
L3

Figure III-26: Stereo views of the antibody CDRs of Fab Hed10. The trace of the polypeptide backbone is approximated by a ribbon and the six CDRs are labelled. **A.)** View from the L3 side of the combining region and **B.)** view looking directly down into the combining region.

III.6 Conclusions

We have approached the question of how antibodies recognize DNA by determining the structures of antibody Fab fragments that bind to specific DNA sequences using X-ray crystallographic techniques. Unfortunately, we are limited in our ability to address this question because we have not obtained any crystallographic data on Fab-DNA complexes. The structures of the Fab fragments themselves may provide some insights into antibody-DNA recognition. Detailed knowledge of the structures of the antibody combining sites, and of the nucleic acids that they bind, may aid in the development of models for their specific interaction. The interpretation of the structural results that we have obtained, however, is limited for the following reasons.

The first limitation is the relatively low resolution of the data. At resolutions between 2.5 and 3.0 Å the interpretation of the electron density can be fraught with errors. Crystal structure analyses at these moderate resolutions have a coordinate error of approximately 0.5 Å for the most accurately determined regions of the molecule and even less accurate coordinates for the poorly determined regions (Davies et al., 1990). These disordered regions are likely to be on the surface of the molecule and exposed to solvent. In Fab fragment structures the conformations of the main chain, and particularly the side chain, atoms of the hypervariable loops are among the least accurately determined portions of the structure. Although these errors may not be readily apparent with data at moderate resolution, when high resolution data (< 2.0 Å) is available the conformations of some of the hypervariable loops are observed to differ significantly from those determined with the lower resolution data.

A rebuilding of all three CDRs of the heavy chain occured when the structural analysis of Fab J539 was extended to high resolution. The initial structure of J539 was determined at a resolution of 2.6 Å (Suh et al., 1986). When 1.95 Å resolution data became available the rebuilding of the heavy chain CDRs resulted in a difference of over 8.0 Å in the coordinates of two corresponding Cα atoms (Davies et al., 1990). A similar situation was observed in the crystal structure of the human immunoglobulin Fab fragment New where the position of some of the residues in CDR H3 was reinterpreted with 2.0 Å data, resulting in an RMS difference of nearly 5.0 Å in the positions of corresponding Cα atoms (Saul and Poljak, 1992). The Fab fragment structure determinations described in this chapter were all performed at relatively moderate resolutions. In light of the aforementioned results these structures must, therefore, be considered as preliminary models for the antibody combining sites.

A second limitation on our ability to address the structural aspects of antibody-DNA recognition is the fact that there may be an "induced fit" of the antibody combining site to the nucleic acid antigen. Significant alterations in the conformations of the hypervariable loops upon antigen binding would make it difficult to predict which residues will interact with specific regions of the nucleic acids based on an analysis of the unliganded Fab fragment structure. Nonetheless, there are several interesting features of the antibody combining sites whose structures we have determined that lead us to speculate on their potential interactions with nucleic acids. These proposed interactions can be tested by site specific mutagenesis and the models for DNA binding by these antibodies refined accordingly. These DNA binding models will be described in Chapter V.

Taking these limitations on our ability to address the precise structural details of antibody-DNA recognition into account, we became intrigued by the possibility that perhaps suitable descriptions of the antibody combining regions could be derived by modelling the structures of the hypervariable loops based on their amino acid sequences and the structures of previously determined antibody Fab fragments. Being in possession of the refined crystal structures of three DNA binding antibody Fab fragments we were, thus, able to assess how similar the properties of the model structures were to the observed crystal structures. The modelling protocol that we have employed is described in the next chapter, along with a comparison of the model and crystal structures.

IV. Modelling of Antibody Combining Regions

IV.1 Introduction

In order to understand how an antibody might interact with its antigen the conformations of the six hypervariable loops must be determined with a reasonable level of accuracy. While the structures of an appreciable number of immunoglobulin Fab fragments have been determined by X-ray crystallography, the sequences of a far greater number of antibodies are known. This fact has led to a number of attempts at predicting antibody loop conformations based on sequence information alone (see below).

Two different predictive methods have been previously employed. The de Novo approach generates loops ab initio based on energetic constraints, while the knowledge or template-based approach uses loops from known crystal structures as models for the new loops. Several groups have reported some success predicting protein loop conformations de Novo (Bruccoleri and Karplus, 1987; Fine et al., 1986; Moult and James, 1986), while others have proven equally as successful using the template-based approach (Chothia et al., 1989; de la Paz et al., 1986; Chothia, et al., 1986; Smith-Gill et al., 1987), and still others have employed a combination of the two approaches (Martin et al., 1989). As the database of medium to high resolution Fab structures grows, it becomes increasingly feasible to model the combining sites of other antibodies based on these known structures using a knowledge based approach. Some of the amino acid sequences of antibodies of unknown structure possess significant homology, or even sequence identity, with known Fab fragment structures. One possible problem with this approach is that the hypervariable loops used as models may come from medium resolution structures and, therefore, not have accurately determined conformations. As more high resolution structures of Fab fragments are determined this potential shortcoming should become less of a problem.

Another potential complication is that the conformations of the antibody CDRs can be influenced by interactions with the framework residues and the other hypervariable regions of the molecule. In order to predict the precise conformation of antibody CDRs it may be necessary to know the exact mode of association of the V_L and V_H domains (Stiepe et al., 1992). However, the purpose of these modelling experiments is not to derive descriptions of the antibody combining regions at the level of precise atomic detail, but rather to examine the general properties of the combining regions of DNA binding antibodies. The distribution of positively charged residues within the combining region may indicate areas of the antibody that are likely to interact with the negatively charged sugar phosphate backbone of the DNA. If the orientation of the DNA backbone with respect to the combining region can be fixed, then antibody residues that are in position to potentially interact with the DNA bases can be inferred. Thus, it may be possible to derive plausible models for specific antibody DNA interactions that can aid in the interpretation of biochemical and immunologic results, and can be experimentally tested and improved upon in the absence of crystallographic data for the Fab and Fab-DNA complexes.

In this chapter the modelling protocol used to predict the loop conformations of six nucleic acid binding antibodies is described. Hed10 is a murine autoimmune antibody specific for single stranded poly(dT) (Lee et al., 1982). Jel 318 is a murine monoclonal antibody that binds triple stranded $\mathrm{poly(dTm^5C)\cdot poly(dGA)\cdot poly(dm^5C^+T)}$ (Lee et al., 1987), while Jel 72 is a mouse monoclonal antibody specific for the right handed DNA duplex poly(dG)·poly(dC) (Lee et al., 1984). The crystal structures of all three Fab fragments have been determined. The level of accuracy of the modelling protocol can be evaluated by comparing the model structures with the structures that have been determined crystallographically.

In addition, models have been constructed of three Fv's for which no crystal structures have yet been determined. Jel 242 is a murine autoimmune antibody that binds double stranded 'B' form DNA (Braun and Lee, 1986), while Jel 201 is a mouse monoclonal antibody specific for poly(ADP-ribose) (Sibley et al., 1986), and Jel 274 is a murine autoimmune antibody that binds d(GC)-rich duplex DNA (J.S. Lee, personal communication).

We have used the canonical structure model for the hypervariable loops, as described by Chothia and Lesk (1987), to derive models for the antibody combining regions. Chothia and Lesk (1987) have examined the known structures of antibody variable domains and have proposed that there is a limited repertoire of main chain conformations for all of the six hypervariable regions of antibodies, with the possible exception of the H3 hypervariable loop. Moreover, they propose that the particular conformation of a hypervariable loop is determined by a small number of conserved residues. If a loop from an antibody of unknown structure is the correct length and these key residues are present in the amino acid sequence, then the loop is assumed to belong to one of the discrete conformational classes, or canonical structures, observed in previously determined antibody Fab fragment structures. The main chain conformation is unaffected by the identity of the other amino acids of the hypervariable loop. Hypervariable regions with similar conformations can, thus, present different surfaces for interaction with antigen by altering the residues that are not involved in stabilizing the loop conformation. The portions of the hypervariable regions that lie outside of the β-sheet framework, and are proposed by Chothia and Lesk to possess structural diversity in different antibodies, are less extensive than the complementarity determining regions defined by Kabat et al. (1987) on the basis of amino acid sequence similarities (see Appendix A).

The variation in the length and sequence of the H3 loop is more extensive than for the other hypervariable loops. For this reason it is difficult to assign the H3 loop to any particular conformational class. Nevertheless, an analysis of the known amino acid sequences from mouse antibodies (Chothia and Lesk, 1987) indicates that many of the H3 loops may be in conformations that are similar to that observed for the H3 loop of the McPC603 Fab fragment (Segal et al., 1974), which forms a long hairpin turn. This similarity was exploited in the successful prediction of the conformation of the H3 loop, as well as the other five hypervariable loops, of the D1.3 anti-lysosyme antibody prior to the determination of the structure of the Fab fragment (Chothia et al., 1986). The modelling protocol proposed by Chothia and Lesk has also been used to predict the hypervariable loop conformations of four other immunoglobulin Fab fragments whose crystal structures were subsequently determined (Chothia et al., 1989), although for these antibodies the structures of the H3 loops could not be predicted. It is proposed that the structures of these loops might be predicted correctly using a conformational search procedure (eg. Moult and James, 1986). The possibility also exists that, as more antibody Fab fragment structures are determined, other discrete conformational classes for the H3 hypervariable loop will become apparent. It is also possible that many of these H3 loops are flexible and do not possess well-defined conformations.

IV.2 Model Building

An examination of the amino acid sequences of the V_H and V_L domains of the nucleic acid binding antibodies revealed that many of the sequences possessed significant similarity to antibodies with known crystal structures. This amino acid sequence similarity is mostly restricted to the V_L domains. Additionally, in most of the sequences examined the hypervariable loops belonged to one of the canonical structures described by Chothia and Lesk (1987). For this reason we decided to attempt to model the combining sites of the nucleic acid binding antibodies using the loop designations and procedures described by them. The sequences of the hypervariable loops are presented in Table IV-1. Also shown are those key residues that determine the loop conformation. The canonical structure class to which each loop was assigned, and the name of the known structure from which the model was derived, is also indicated in the Table IV-1. In the case of the L1 loop from Hed10 coordinates for the indicated canonical structure were not available from the Brookhaven Protein Data Bank. The model for this loop was constructed from the L1 loop of McPC603 by excising residue 31B and rotating the newly exposed end residues to bring them within peptide bonding distance. For the L1 loops of Jel 72 and Jel 201, which belong to the same canonical structure class as Hed10, the model loop was constructed using the L1 loop from the Hed10 crystal structure (Cygler et al., 1987).

Table IV-1: A.) Amino Acid Sequences for the Hypervariable Loops of the V_L domains of nucleic acid binding Antibodies

L1

						*								*	*	*	*	
CS	Prot.	26	27	a	b	c	d	e	28	29	30	31	32	2	25	33	71	Model
4	Hed10	S	T	S	L	L	H	S	S	G	K	N	R	I	S	L	F	4-4-20
4	JEL 72	S	Q	S	L	L	D	S	D	G	K	T	Y	V	S	L	F	Hed10
1	JEL 318	S	-	-	-	-	-	-	S	S	V	S	Y	I	A	M	Y	J539
2	JEL 242	S	Q	-	-	-	-	-	D	I	S	N	N	I	A	L	Y	HyH-10
4	JEL 201	S	K	S	L	L	H	S	N	G	N	T	Y	I	S	L	F	Hed10
4	JEL 274	S	Q	S	L	L	Y	S	N	G	K	T	Y	V	S	L	F	Hed10

L2

					*	*	
CS	Protein	50	51	52	48	64	Model
1	Hed10	Y	M	S	I	G	J539
1	JEL 72	L	V	S	I	G	J539
1	JEL 318	D	T	S	I	G	McPC603
1	JEL 242	Y	T	S	I	G	HyHEL-10
1	JEL 201	R	M	S	I	G	J539
1	JEL 274	L	V	S	I	G	JEL 72

L3

						*		*	
CS	Protein	91	92	93	94	95	96	90	Model
1	HED10	S	L	Q	Y	P	Y	Q	McPC603
1	JEL 72	G	T	H	F	P	Q	Q	McPC603
1	JEL 318	L	S	S	N	P	Y	Q	McPC603
1	JEL 242	G	N	T	L	P	R	E	HyHEL-10
1	JEL 201	H	L	E	Y	P	Y	Q	McPC603
1	JEL 274	G	T	H	F	P	Y	Q	JEL 72

Loops were constructed with the following coordinates taken from the Brookhaven Protein Data Bank (Bernstein et al., 1977): PDB1MCP.ENT, McPC603 (Satow et al., 1987); PDB2FBJ.ENT, J539(Suh et al., 1986); PDB2HFL.ENT, HyHEL-5 (Sheriff et al., 1987); PDB3HFM.ENT, HyHEL-10 (Padlan et al.,1989); PDB4FAB.ENT, 4-4-20 (Herron et al., 1989).

Table IV-1: B.) Amino Acid Sequences for the Hypervariable Loops of the V_H domains of nucleic acid binding Antibodies

H1

		*	*		*				*	*	
CS	Protein	26	27	28	29	30	31	32	34	94	Model
1	HED10	G	F	S	F	S	N	Y	M	V	McPC603
1	JEL 72	G	Y	T	F	T	S	Y	M	K	McPC603
1	JEL 318	G	F	S	F	N	T	Y	M	R	McPC603
1	JEL 242	G	Y	T	F	T	S	Y	I	R	J539
1	JEL 201	A	F	T	F	S	D	Y	M	R	McPC603
1	JEL 274	G	Y	T	F	T	N	S	I	R	J539

H2

							*	*	
CS	Protein	52a	52b	52c	53	54	55	71	Model
4	HED10	L	R	S	D	N	Y	R	McPC603
2	JEL 72	P	-	-	R	S	G	V	HyHEL-5
4	JEL 318	S	K	S	K	N	Y	R	McPC603
2	JEL 242	P	-	-	R	S	G	V	J539
2	JEL 201	S	-	-	G	S	T	R	J539
2	JEL 274	P	-	-	G	S	S	V	J539

H3

	*												*	
Protein	94	95	96	97	98	99	100	a	b	c	d	e	101	Model
HED10	V	Y	Y	G	Y	L	-	-	-	-	-	-	G	HyHEL-5
JEL 72	K	G	G	S	R	V	R	R	Y	Y	A	M	D	McPC603
JEL 318	R	E	L	L	R	S	F	-	-	-	-	-	A	J539
JEL 242	R	Y	G	N	Y	V	R	F	-	-	-	-	D	J539
JEL 201	R	N	N	Y	Y	G	S	S	P	F	-	-	A	McPC603
JEL 274	R	W	R	D	Y	R	S	F	F	-	-	-	A	McPC603

Loops were constructed with the following coordinates taken from the Brookhaven Protein Data Bank (Bernstein et al., 1977): PDB1MCP.ENT, McPC603 (Satow et al., 1987); PDB2FBJ.ENT, J539(Suh et al., 1986); PDB2HFL.ENT, HyHEL-5 (Sheriff et al., 1987); PDB3HFM.ENT, HyHEL-10 (Padlan et al.,1989); PDB4FAB.ENT, 4-4-20 (Herron et al., 1989).

Parent structures for the V_H and V_L framework regions were selected and those amino acids that differed were replaced with the amino acids corresponding to the nucleic acid binding antibodies in such a way as to maximize side chain overlap. Model loops were grafted onto these framework structures by superimposing the backbone atoms (N,C,Cα) of the three residues preceding and the three residues following the loop onto the corresponding atoms of the parent domains. For the special case of the H3 loop, models were selected based on their length and sequence similarity. If an existing H3 loop of the same size was not available then the nearest length loop was selected and the residue(s) that were deemed to be at the apex of the loop were added or excised. The loop was closed by rotating the two newly exposed end residues about their phi and psi torsion angles to bring the respective nitrogen and carbonyl carbon atoms within approximate peptide bonding distance. Sequence considerations were based almost exclusively on whether both the modelled and existing loops contained an arginine at position H94 and an aspartic acid at position H101. In the McPC603 Fab structure these two residues form a salt link at the base of the H3 loop (Satow et al., 1987).

After the six loops had been assembled, the V_H and V_L domains of the model structures were superimposed onto the V_H and V_L domains of McPC603 by the backbone atoms of the conserved framework residues (V_L: 4-6, 9, 11-13, 19-25, 33-49, 53-55, 61-76, 84-90, 97-106; V_H: 3-12, 17-25, 33-52, 56-60, 68-82, 88-95, 102-112 (Getzoff et al., 1988)). These residues are considered to be structurally conserved because the differences in the positions of atoms in homologous residues is small when the main chain atoms of all known Fab fragment structures are superimposed (Chothia and Lesk, 1987). Thus, all of the modelled structures possessed the variable domain association as found in McPC603. The pseudo-twofold rotation relating the V_H and V_L domains in McPC603 lies approximately midway between the extremes observed in known crystal structures of immunoglobulin Fab fragments.

The resulting model Fv structures were then energy minimized using the Powell-method of conjugate gradient minimization (Powell, 1971) as implemented in X-PLOR. The model Fv structures often possessed unfavourable van der Waals energies resulting from the replacement of small side chain amino acids with amino acids that had larger side chains. A very limited amount of energy minimization was required to correct for these unfavourable contacts, usually from 15 to 25 cycles of minimization. The resulting differences in the minimized and unminimized structures were relatively small, with movements of the side chain atoms of some residues on the order of 0.2 to 0.4 Å.

IV.3 Comparison of Model and Crystal Structures

These model structures were then compared with the crystal structures to determine the degree of accuracy of the modeling protocol. The results indicate that where the model and crystal structures differ can be attributed, at least in part, to crystal packing interactions.

The modeled structures of the hypervariable loops are compared to the crystallographically determined conformations in Figure IV-1. Root mean square (RMS) deviations from aligned positions for the backbone atoms of both the conserved β-sheet framework residues, and the hypervariable loops, as well as the RMS deviations of all the atoms in the loops, are listed in Table IV-2. The positions of the backbone atoms of all of the modeled loops, with the exception of H3, are within 1.0 Å deviation of the crystallographically observed loop conformations. The backbone atoms of the H3 loops deviate by only slightly more than 1.0 Å (Hed10) to just over 2.5 Å (Jel 72). When all of the side chain atoms are included the RMS deviation for most of the loops increases to between 1.0 and 2.0 Å. The particularly large deviations of the side chain atoms for the H3 loops and the H2 loop of Jel 318 are mostly attributable to radically different orientations of a few residues with relatively large side chains (lys, arg, tyr). The problematic residues are, for the most part, exposed to solvent, or at least would be in a free Fab fragment, and may exist in multiple conformations.

A.

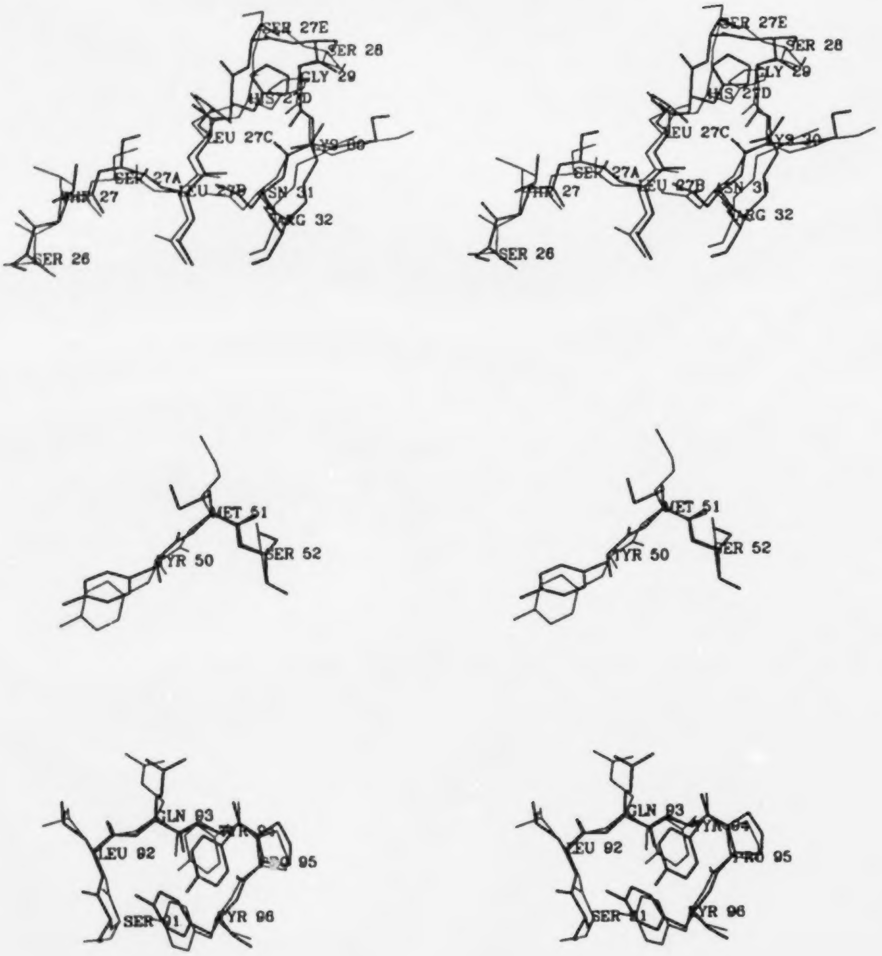

SER 27E
SER 28
GLY 29
HIS 27D
LEU 27C
LYS 30
SER 27A
THR 27
LEU 27B
ASN 31
ARG 32
SER 26
MET 51
TYR 50
SER 52
GLN 93
LEU 92
TYR 94
PRO 95
SER 91
TYR 96

Figure IV-1: Hypervariable loop conformations. Stereo representations of the conformations of the hypervariable loops observed in the model (thick lines) and crystal (thin lines) structures. The loops have been superimposed according to a least squares fit of their backbone atoms (N, C, Cα) **A.)** Hed10, Light chain Hypervariable regions: L1 (top), L2 (middle), and L3 (bottom).

B.

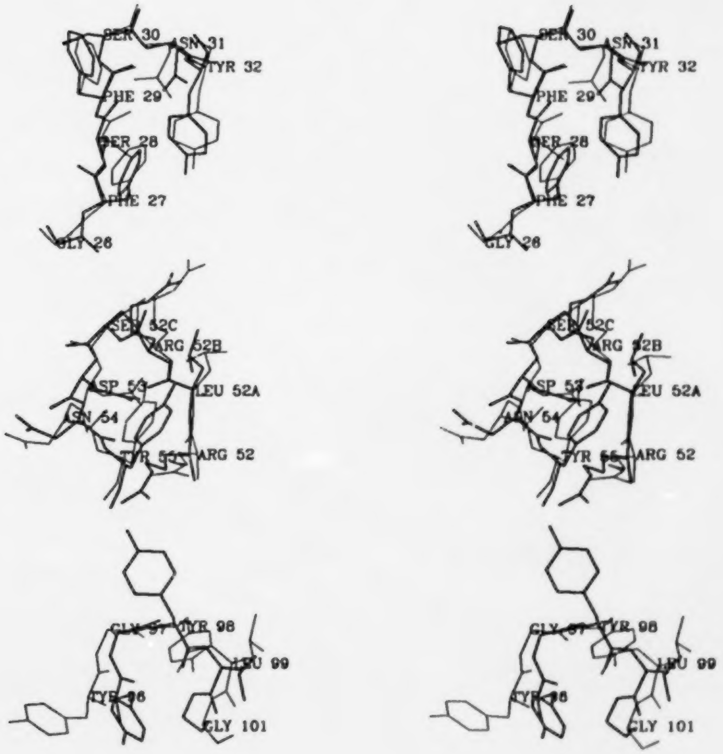

SER 30
ASN 31
TYR 32
PHE 29
SER 28
PHE 27
GLY 26
SER 52C
ARG 52B
ASP 53
LEU 52A
ASN 54
TYR 55
ARG 52
GLY 97
TYR 98
LEU 99
TYR 96
GLY 101

Figure IV-1: Hypervariable loop conformations. **B.)** Hed10, Heavy chain Hypervariable regions: H1 (top), H2 (middle), and H3 (bottom).

c.

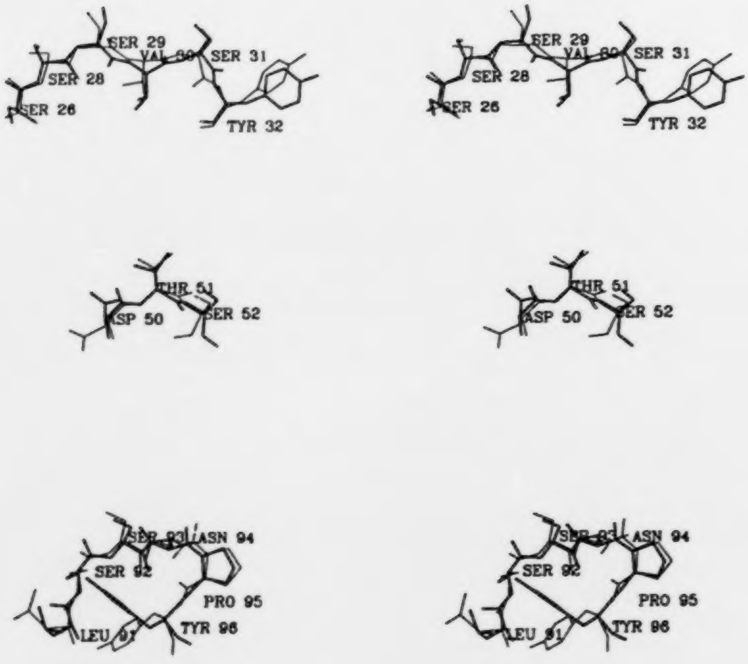

SER 29
VAL 30
SER 31
SER 28
SER 26
TYR 32
THR 51
ASP 50
SER 52
SER 93
ASN 94
SER 92
PRO 95
LEU 91
TYR 96

Figure IV-1: Hypervariable loop conformations. **C.)** Jel 318, Light Chain Hypervariable regions: L1 (top), L2 (middle), and L3 (bottom).

D.

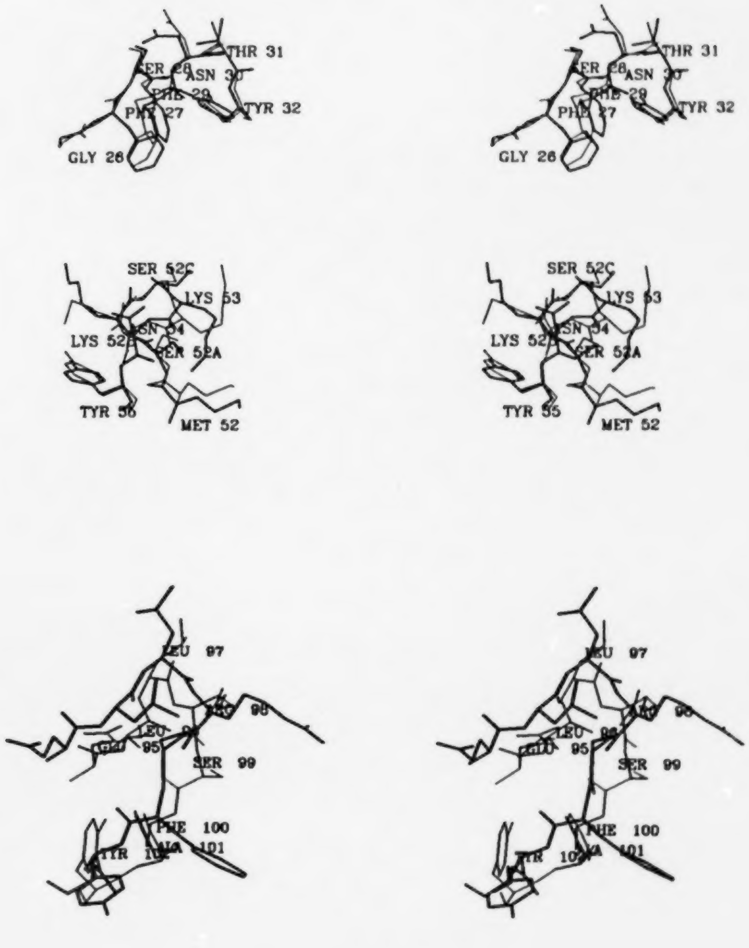

THR 31
SER 28
ASN 30
PHE 29
PHE 27
TYR 32
GLY 26
SER 52C
LYS 53
LYS 52B
ASN 54
SER 52A
TYR 56
MET 52
LEU 97
98
LEU 96
GLU 95
SER 99
PHE 100
ALA 101
TYR 102

Figure IV-1: Hypervariable loop conformations. **D.)** Jel 318, Heavy Chain Hypervariable regions: H1 (top), H2 (middle), and H3 (bottom).

E.

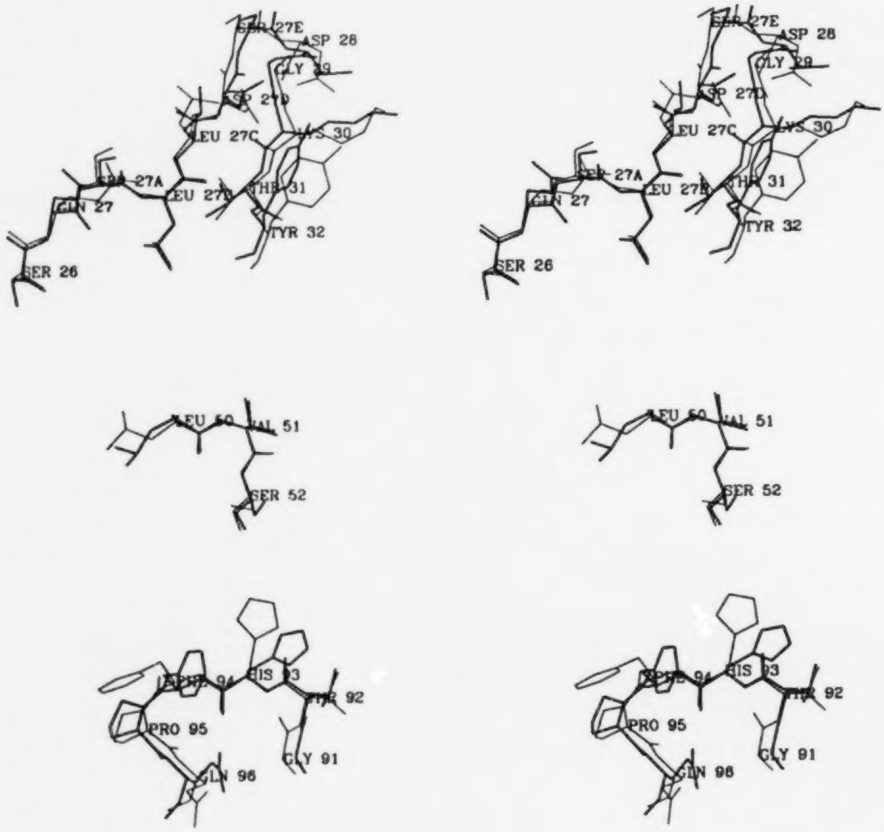

SER 27E
ASP 28
GLY 29
ASP 27D
LEU 27C
LYS 30
SER 27A
GLN 27
LEU 27B
THR 31
TYR 32
SER 26
LEU 50
VAL 51
SER 52
PHE 94
HIS 93
THR 92
PRO 95
GLY 91
GLN 96

Figure IV-1: Hypervariable loop conformations. **E.)** Jel 72, Light Chain Hypervariable regions: L1 (top), L2 (middle), and L3 (bottom).

F.

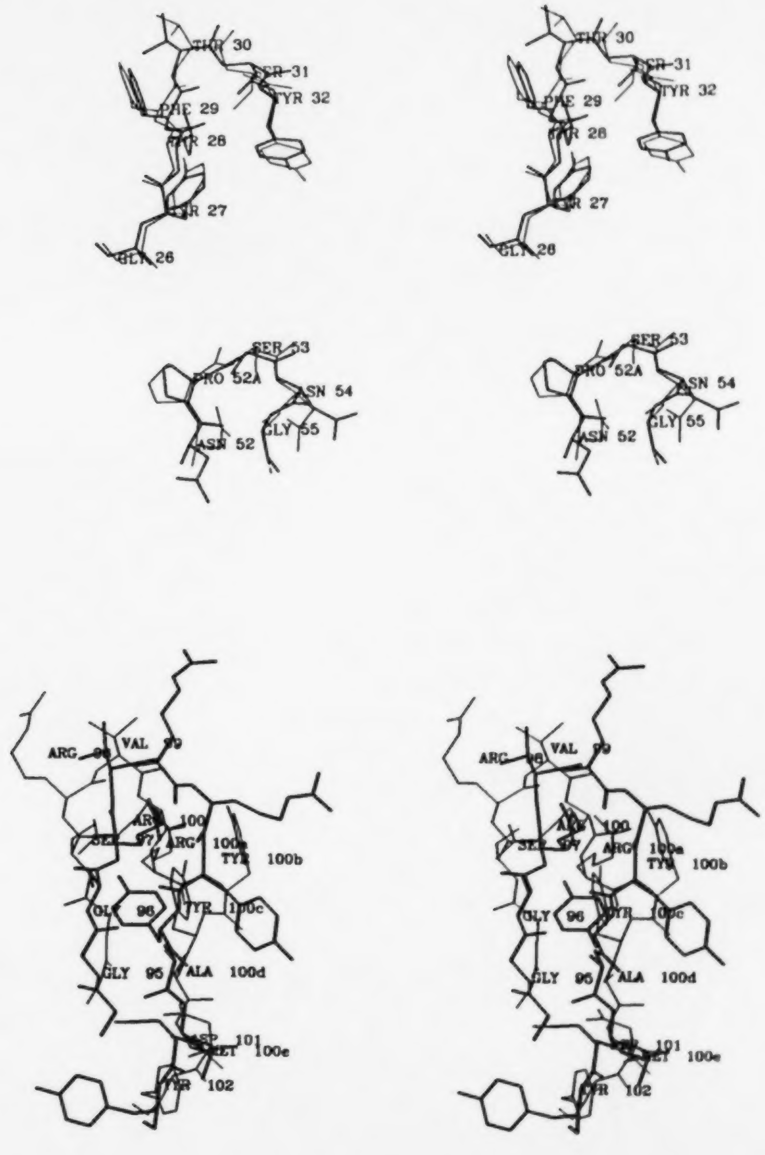

THR 30
SER 31
TYR 32
PHE 29
TYR 28
TYR 27
GLY 26
SER 53
PRO 52A
ASN 54
GLY 55
ASN 52
ARG 98
VAL 99
ARG 100
SER 97
ARG 100a
TYR 100b
GLY 96
TYR 100c
GLY 95
ALA 100d
ASP 101
MET 100e
TYR 102

Figure IV-1: Hypervariable loop conformations. **F.)** Jel 72, Heavy Chain Hypervariable regions: H1 (top), H2 (middle), and H3 (bottom).

Table IV-2: RMS Deviations from aligned positions of modeled loops

	Hed10		Jel 72		Jel 318	
	#Atoms	RMS	#Atoms	RMS	#Atoms	RMS
V_L	201	0.64	201	0.61	201	0.77
L1	36	0.79	36	0.62	18	0.65
	89	1.24	91	1.24	43	1.20
L2	9	0.53	9	0.16	9	0.36
	26	1.61	21	0.84	21	1.30
L3	18	0.35	18	0.47	18	0.38
	54	1.07	48	1.69	47	1.69
V_H	234	1.15	234	0.73	234	0.56
H1	21	0.33	21	0.49	21	0.37
	58	0.69	59	0.72	59	0.63
H2	21	0.55	15	0.49	21	0.69
	64	1.11	33	1.66	58	1.78
H3	21	0.99	39	2.64	24	1.56
	63	2.89	111	5.91	70	3.50

The first row of numbers refers to the backbone atoms (N,C,C_α), the second row of numbers is RMS deviations for all non-hydrogen atoms.

V_L: RMS deviations for the backbone atoms of conserved framework residues (4-6,9,11-13,19-25,33-49,53-55,61-76,84-90,97-106).

V_H: RMS deviations for the backbone atoms of conserved framework residues (3-12,17-25,33-52,56-60,68-82,88-95,102-112), using the residue numbering of Kabat et al.(25).

In fact, the conformations of at least some of these side chains with different orientations may be at least partially influenced by crystal packing interactions. In crystals of Fab fragments, they Fabs are often observed to pack tightly against symmetry related molecules. The variable domains of both Hed10 and Jel 318 pack closely against constant domains of symmetry related molecules, while in Jel 72 the packing involves interactions between two variable domains related by non-crystallographic symmetry.

Jel 318: The combining site of Jel 318 is a relatively flat surface punctuated by the protrusion of a knob comprised of the H2 loop on the V_H side of the combining region. The crystal packing of Jel 318 involves contacts made almost exclusively by the H2 loop with the COOH terminus of a symmetry related C_L domain. The conformation of lysine 53 is influenced by these interactions, with a possible hydrogen bond existing between the terminal nitrogen on the lysine side chain and the carbonyl oxygen of C_L serine 207 (Figure III-17). The conformation of the side chain of this lysine differs markedly in the modeled loop (Figure IV-1d), and it is this large difference in overall side chain orientation that contributes to the unusually large RMS deviation of all of the atoms in this loop.

Hed10: The combining region of Hed10 is characterized by a deep cleft bordered on the V_L side by the L1 loop and on the V_H side by the H2 loop. Despite the fact that the global packing arrangement of Hed10 is similar to Jel 318, the specific crystal contacts of Hed10 differ markedly from those of Jel 318. Most of the crystal contacts in Hed10 involve residues from the V_L domain interacting with a symmetry related C_L chain. These contacts are dominated by residues of the L1 loop. Interestingly, while the conformation of the L1 loop in Hed10 was modelled satisfactorily, the disposition of this loop with respect to the conserved framework residues of the V_L domain was different. When the modeled V_H and V_L domains are superimposed onto the crystal structures by their conserved framework residues the differences in the positions of the α-carbons of the loops are indicative of whether the entire loop is shifted in its position relative to the other loops. Results from this type of analysis indicate that the L1 loop of Hed10 is moved inward, toward the combining site, by nearly 3.5 Å, in the crystal structure relative to the modeled structure. This positioning of the L1 loop in the crystal allows it to make several crystal contacts and potential hydrogen bonds, specifically between L1 histidine L27d and C_L glutamine 218. The relative position of any modeled loop could just as easily be influenced by the crystal packing interactions in the structure from which the loop, or parent framework domain, was derived.

Jel 72: The combining region of Jel 72 forms a slightly concave surface with residues of the L1 loop forming a ridge on the V_L side of the combining region and residues in the H2 loop forming a smaller ridge on the V_H side, with the long arginine rich H3 loop occupying a central position in the combining region. Jel 72 crystallizes in a rather atypical fashion with two variable domains related by a non-crystallographic two-fold axis packing against one another. This results in residues of CDR H2 packing against residues within L1 and L3 of the other variable domain. This interaction is asymmetric. The side chain conformations of some of the residues in H2, particularly lysine H64, differ in the two variable domains, such that the modeled H2 loop more closely resembles the loop as seen in Fab two of Jel 72 in crystal form I. The asymmetry of the crystal packing is reflected in the binding of the heavy atom derivative $Pt[(CNS)_6]^{2-}$. This heavy atom binds to histidine L93 in L3 of one V_L domain but not to the other.

If these crystal packing artifacts are taken into account, satisfactory models have been derived for all of the loops with the possible exception of H3. Considering that the modeling was purposefully kept rather crude, with no attempt being made to determine preferred side chain orientations, even the H3 loop results are not discouraging. The H3 loop of Jel 72 contains the sequence GLY-GLY-SER which is where the electron density begins to become difficult to interpret, especially for the side chains of the arginine residues. This can be an indication of either discrete or dynamic disorder of this portion of the H3 loop in the crystal. For similar long, glycine containing, H3 loops in other antibodies there may not be a single preferred conformation, but instead the loop is relatively flexible and assumes a conformation to maximize interaction with the ligand during binding. It has been noted in the co-crystal structure of the Fab BV04-01 and the trinucleotide d(pT_3) that the residues of the H3 loop have higher temperature factors in the unliganded Fab than in the Fab-trinucleotide complex (Herron et al., 1991). By a comparison of the structures of the complexed and uncomplexed Fabs they also note a significant conformational change of the H3 loop upon binding the DNA. This observation reinforces the point that for antibodies in general, and DNA binding antibodies in particular, conformational flexibility of the hypervariable loops, particularly H3, may be an important factor in maximizing the complementarity between the nucleic acid antigen and the antibody combining site.

IV.4 Discussion

There are several reasons why we have attempted to model the combining regions of DNA binding antibodies. The crystal structure determinations of all three antibody Fab fragments; Hed10 (Cygler et al., 1987), Jel 72 and Jel 318, were hindered by the lack of amino acid sequence information for their variable domains. Once the amino acid sequences were determined the structural analyses were able to proceed. Even without this impediment, however, a crystal structure determination can still consume significant resources, take several months to complete, and is, of course, contingent on the fact that suitable crystals have first been obtained. The results presented in Chapter II illustrate that obtaining crystals of antibody Fab fragments is a relatively straightforward, although by no means trivial, procedure.

A model for the antibody combining site, however, can be derived in much less time than is required for a crystal structure determination. Since the crystal structures of Hed10, Jel 72 and Jel 318 were already being determined, a major motivation for modeling their combining regions was to evaluate whether reasonable model structures could have been derived had the amino acid sequences of their variable regions been available at an earlier date. A definition of what is meant by 'reasonable' can be obtained by comparing the data presented in Table IV-2 to the estimated mean coordinate error of the crystal structures.

The RMS deviations of the modeled loops from the crystallographically determined loop conformations are within approximately 1.0 Å for the backbone atoms, and, generally within 2 to 3 Å when side chain atoms are included in the analyses. The estimated mean coordinate error of the crystal structures is on the order of 0.5 Å, but this value might even be higher for the poorly determined, solvent exposed loops of the antibody combining region, particularly for the precise details of the side chain torsion angles. Thus, the modeled loops represent structures that are roughly equivalent to the crystal structures at the level of the relative disposition of the hypervariable loops with respect to one another.

As mentioned in Chapter III, the interpretation of the crystal structures was limited by the lack of high resolution data and the apparent flexibility of some of the hypervariable regions. The conformations of the hypervariable loops can also be influenced by crystal packing interactions. Similar situations exist for many of the structures from which the models were constructed. The structure of the McPC603 Fab fragment, that served as a model for many of the hypervariable loops, has been determined at moderate resolution (Satow et al.1987), as are many of the crystal structures of Fab fragments. The database of Fab crystal structures that were employed to derive the models was rather small considering the number of Fab fragment structures that have been determined in recent years. As more, and particularly higher resolution, structures are determined (see eg. Saul and Poljak, 1992; and Stiepe et al., 1992) the accuracy of the model structures can be expected to improve. The exploitation of this facility to derive 'reasonable' models of antibody combining sites will depend on a careful and rational application of the modeling protocol. In this instance, Hed10, Jel 72 and Jel 318 were all excellent candidates to have their combining sites modeled prior to having their structures determined. All three antibodies bind to specific DNA sequences or structures, whereas Jel 242 and other autoimmune double stranded DNA binding antibodies (Shlomchik et al., 1987) bind 'B'-form DNA with little base sequence preference (Braun and Lee, 1986). If the modeled combining regions possess interesting features then the decision can be made on whether to endeavour to determine the crystal structures of the Fab fragments.

The polyanionic nature of nucleic acids makes it likely that electrostatics will play an important role in antibody-nucleic acid recognition. We were, thus, interested in determining whether the model combining regions possessed potential energy surfaces similar to the crystal structures. An examination of the surface potentials of the combining sites of the crystal structures has aided us in deriving testable models for DNA binding by these antibodies. These DNA binding models will be discussed in detail in the next chapter.

The electrostatic surface potential for Hed10 is compared with the potential of the model for Hed10 in Figure IV-2. Electrostatic surface potentials for both the modeled and observed structures were calculated using the DELPHI program (Biosym Technologies, Inc.). The solute dielectric was set to 2.0 and the solvent dielectric to 80.0. The ionic radius of the solvent was 1.40 Å and an ionic strength of 0.145 M was used. The potential energy surface of the molecules is dominated by the negative charges on aspartic and glutamic acid and the positive charges of lysine and arginine side chains. For this reason, and because our interest lies in examining the general features of nucleic acid binding antibody combining regions, only the protein formal charges were used in the calculations. The view in the figure is along the cleft of the combining site. The interior of the cleft is lined with many aromatic amino acid side chains, mostly tyrosines of the H3 loop, while on the sides of the cleft there are regions of positive electrostatic potential. These features are also apparent in the model for the combining site of Hed10. We have observed this general correlation to also be true for Jel 72 and Jel 318. Although the potential surfaces of the modeled structures are less focussed and discrete than the crystal structures, the general features of their combining sites do correspond with the observed potentials. Thus, conclusions about the orientation of the DNA antigen derived from model structures may provide useful information for interpreting biochemical and immunologic data.

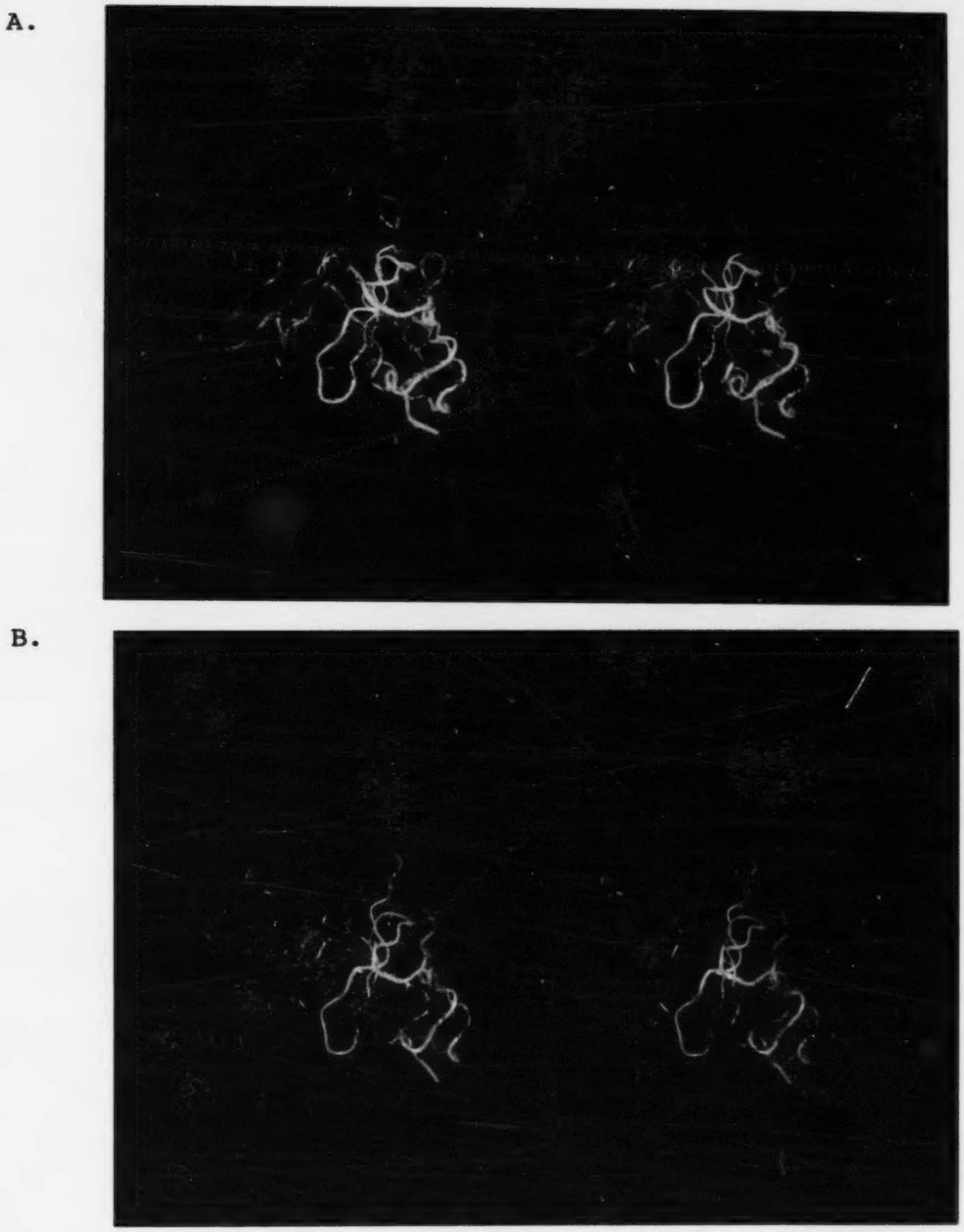


Figure IV-2: Electrostatic Surface Potentials for modeled and observed Fv domains a) Hed10 crystal structure, b) Hed10 model structure.

V. DNA Binding Models

V.I Introduction

An examination of the shape and electrostatic surface potential of the crystallographically determined Fab fragment structures has aided us in deriving testable hypotheses for how these antibodies might interact with their respective nucleic acid antigens. The purpose of these DNA binding models is to provide a structural framework that can assist in the interpretation of solution data. It must be reiterated that there is presently no crystallographic data for the Fab-DNA complexes that can be used to confirm or refute the binding mechanisms that will be proposed. It is also unclear how the impressive amount of structural data that has been obtained on antibody binding to protein antigens can be applied to antibody nucleic acid interactions.

The available structural and biochemical data on antibody binding to protein antigens has been analyzed by Getzoff and co-workers (1988) and the following equation for the interaction of antibody (Ab) with protein antigens (Ag) proposed:

$$\mathrm{Ab} + \mathrm{Ag} \underset{\mathbf{k}_2}{\overset{\mathbf{k}_1}{<\!\!-\!\!->}} \mathrm{Ab}\cdots\mathrm{Ag} \underset{\mathbf{k}_4}{\overset{\mathbf{k}_3}{<\!\!-\!\!->}} \mathrm{Ab}'\cdot\mathrm{Ag}'$$

The interaction is, thus, divided into recognition and binding components. The recognition process involves the formation of a pre-association complex (Ab···Ag) through relatively long range electrostatic and polar interactions. The subsequent exclusion of bound water molecules and the potential rearrangement of contact residues in either the antibody or the antigen, or both, leads to the formation of the final antibody-antigen complex (Ab'· Ag'). Antibody binding to nucleic acids may proceed through a similar mechanism, although the relative rates of formation of the pre-association and final complexes are likely to differ. For nucleic acid binding antibodies that possess significant positive electrostatic potential, the rate of formation of the pre-association complex may be quite rapid ($k_1 > k_2$). The formation of a final complex may depend on the ability of the antibody to make base specific contacts with the DNA. For many of the autoimmune DNA binding antibodies that do not seem to have base sequence specificity this step may be prohibitively slow ($k_4 >> k_3$). Whether or not a nucleic acid binding antibody is recognizing specific DNA sequences, the electrostatic component of the interaction is likely to be significant.

An examination of the electrostatic surface potential of antibody combining sites can be useful in determining where charged groups on the antigen may bind. For DNA binding antibodies the primary concern is where the phosphate groups of the DNA backbone will interact. For Jel 72, which binds double stranded 'A'-form DNA, and Jel 318, which binds triplex DNA, the precise spatial arrangement of the phosphate groups on the antigen is known. Thus, the derivation of a binding model is simplified to fixing two points of the DNA helix onto the antibody combining site. For Hed10, which binds single stranded DNA, a precise structure of the nucleic acid antigen can not be assumed. Luckily, a co-crystal structure of the Fab fragment BV04-01 and the DNA $d(pT_3)$ has recently been determined. (Herron et al., 1991). This antibody possesses an affinity for single stranded poly(dT) that is similar to Hed10. DNA binding to BV04-01 is relatively insensitive to ionic strength (Ballard et al., 1985), suggesting that charge neutralization by BV04-01 of the phosphate groups on the DNA is not a major determinant for the formation of stable antibody-DNA complexes. The binding of poly(dT) by Hed10, however, is sensitive to ionic strength. The binding data suggests that Hed10 interacts with two phosphate groups of the DNA (Lee et al., 1982).

By a comparison of the structure of BV04-01 in the Fab-DNA complex with the structure of the unliganded Fab, Herron and colleagues (1991) found results in agreement with the proposal of Colman (1988). Colman proposed that the relative disposition of the V_H and V_L domains can be rearranged upon antigen binding to maximize the complementarity of the antibody-antigen interface. This rearrangement of the quaternary association of the two domains was overshadowed, however, by a significant alteration in the tertiary structure of the hypervariable loops. A large RMS displacement of about 3.0 Å for residues in L1 and H3 was observed relative to the structure of the Fab without bound DNA.

These results are strongly indicative of an induced fit mode of binding. For this reason any model for antibody DNA interaction based on the structures of uncomplexed Fab fragments is bound to be simplistic. The models that are presented here are not intended to be definitive but are meant to serve as a guide in directing future biochemical and immunological experiments.

V.2 Single Stranded DNA

In trying to derive a model for how an antibody might interact with a single stranded nucleic acid one is presented with the problem that there is no clear picture of exactly what conformation the DNA is adopting when bound to the protein. It is well known that single stranded DNA is flexible and can assume many conformations that are influenced by such things as base stacking interactions. Add to this the potential for conformational change on the part of the antibody and the situation becomes extremely complex. A further complication arises from the fact that the antibody may not be recognizing different portions of the DNA molecule equally. The antibody may be interacting in a non-specific manner only with the phosphodiester backbone of the DNA, or it may be recognizing only the DNA bases and have minimal interactions with the DNA backbone.

In the co-crystal structure of BV04-01 and $d(pT_3)$ the antibody interacts with both the bases and the DNA backbone. The phosphates of the trinucleotide are all arrayed on one side of the combining site, but only the phosphate of the second nucleotide participates in a direct charge-charge interaction with an arginine in the H2 loop. The combining region of BV04-01 forms a long cleft into which the thymine bases penetrate. The central thymine of the trinucleotide is held in place by stacking interactions with a tryptophan of the H3 loop and a tyrosine of the L1 loop. The N3 and O4 atoms of this thymine are within hydrogen bonding distance of the backbone carbonyl oxygen and side chain hydroxyl group, respectively, of a serine in the L3 loop (see page 29). The authors suggest that this potential to form double hydrogen bonds contributes to the immunodominance of thymine in many of the autoantigens found in systemic lupus erythematosus.

The combining region of Hed10 is similar to BV04-01. It also forms a long cleft roughly parallel to the V_H-V_L domain interface (Figure V-1). The interior of this cleft is particularly enriched in tyrosines, including three tyrosines from the H3 loop, and is bordered on the V_L side by a large ridge possessing positive electrostatic potential created by residues Lys L30 and Arg L32, and by Arg H52b. The binding of Hed10 to poly(dT) has been examined in solution and it was suggested that two phosphates were interacting with the antibody (Lee et al., 1982). The tyrosine of the L1 loop in BV04-01 that stacks with the central thymine base in the co-crystal structure is replaced by arginine L32 in Hed10. This arginine in Hed10 may be interacting with the phosphates rather than the bases resulting in most of the protein contacts to the thymine bases being mediated by the tyrosines of the H3 loop. The hydrophobic nature of the first three atoms of the arginine side chain, however, does not preclude this residue from stacking with the thymine base in a similar manner as observed between the tyrosine side chain and the thymine base in the BV04-01 d(pT_3) complex structure.

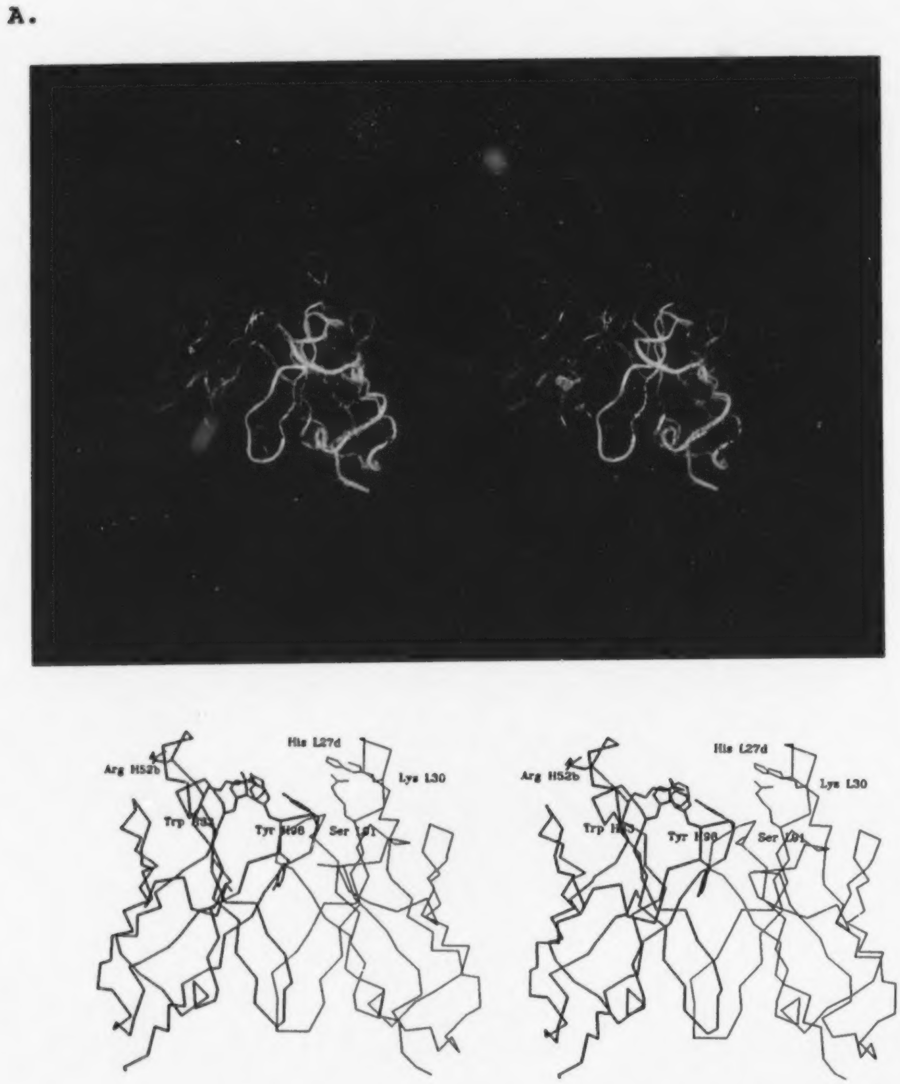

A.
Arg H52b
His L27d
Lys L30
Trp H33
Tyr H98
Ser L91
Arg H52b
His L27d
Lys L30
Trp H33
Tyr H98
Ser L91

Figure V-1: Combining Region and Electrostatic Surface Potential of Hed10. **A.)** View from the H3 side of the combining region. The V_H domain is shown in green (top) and in thick lines (bottom), the V_L domain in yellow (top) and in thin lines (bottom). The trace of the polypeptide backbone is shown as a solid ribbon (top) or as only the Cα atoms (bottom). Surface potentials are shown at contour levels of +1 kcal/mole (red) and -1 kcal/mole (blue), and were calculated as described in Chapter IV.

B.

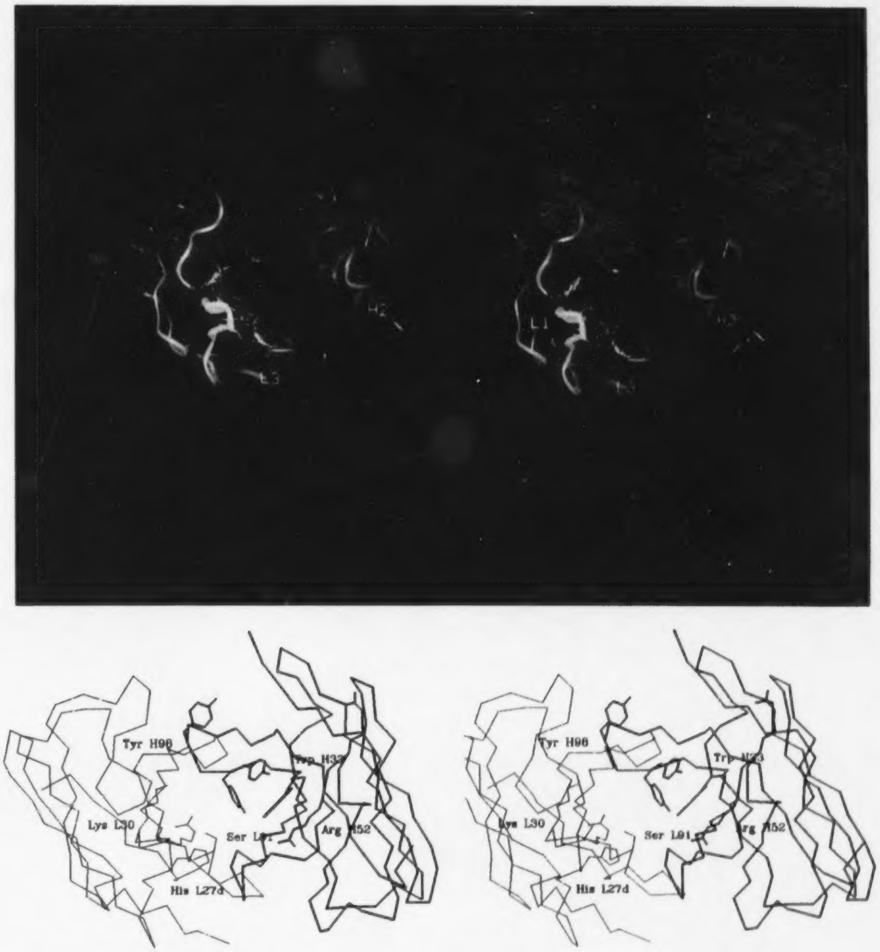

L1
H2
L3
H3
Tyr H98
Trp H33
Lys L30
Ser L91
Arg H52
His L27d

Figure V-1: Combining Region and Electrostatic Surface Potential of Hed10. **B.)** View looking directly down into the combining region. The V_H domain is shown in green (top) and in thick lines (bottom), the V_L domain in yellow (top) and in thin lines (bottom). The trace of the polypeptide backbone is shown as a solid ribbon (top) or as only Cα atoms (bottom). Surface potentials are shown at contour levels of +1 kcal/mole (red) and -1 kcal/mole (blue), and were calculated as described in Chapter IV.

If arginine L32 does, in fact, interact with the DNA backbone, then replacing it with a tyrosine may diminish the effect of ionic strength on the binding of poly(dT) to Hed10, while maintaining or enhancing its affinity for this DNA. In the BV04-01 DNA complex only one arginine in the H2 loop is directly interacting with one of the charged phosphate groups on the DNA. In Hed10 there is an arginine in a similar position, H52b, that may also interact with a phosphate group. Thus, the arginines at L32 and H52b may be contacting two separate phosphates in accordance with the solution data. The current model has the DNA in an extended conformation and places the sugar phosphate backbone along the positively charged ridge on the V_L side of cleft while the thymine bases penetrate into the cleft and stack with the side chains of the tyrosine residues. It is possible that significant conformational changes of the hypervariable loops will accompany DNA binding making it difficult to predict which residues may be involved in base specific contacts. Nevertheless, the potential for forming double hydrogen bonds with the N3 and O4 atoms of the thymines, as observed in the BV04-01 complex, might still be mediated through serine L91. Of the six nucleic acid binding antibodies that we have examined only Hed10 has a serine in this position. It is this serine in BV04-01 that forms the double hydrogen bonds with the thymine base. Other similarities between the two antibodies include a conserved histidine at position 27d of the L1 loop. This histidine forms a hydrogen bond with O3' of the deoxyribose of the second nucleotide in the BV04-01 complex. The thymine base of the third nucleotide was associated with another histidine, L93, as well as threonine L92 and valine L94. In Hed10 residue L93 is a glutamine while L94 is a tyrosine. The H3 loops of the two antibodies are also quite different. Hed10 possesses a shorter H3 loop that is particularly tyrosine rich, while the H3 loop in BV04-01 is longer and contains a single tryptophan that stacks with the DNA base.

Despite these differences, there are many interesting similarities between the combining sites of the two antibodies. A comparison of the structure of Hed10 with both the uncomplexed and complexed structures of BV04-01 should help to clarify some of the details of their interaction with poly(dT). Such an analysis awaits the refinement of the structure of Hed10 with higher resolution data (M. Cygler unpublished results.)

V.3 Double Stranded DNA

The combining regions of the double strand specific anti-DNA antibodies share several features that distinguish them from the single strand specific antibodies. Their combining regions do not form long deep clefts, but instead present relatively flat surfaces, that can be bordered by knoblike ridges consisting of residues in CDR H2 on the V_H side or residues in CDR L1 on the V_L side. These ridges are often the site of strongly positive electrostatic potential, while there is a region of negative potential in the interior of the combining site.

Studies on several autoimmune anti-DNA antibodies have found a correlation between increased affinity for double stranded DNA and the somatic mutation of hypervariable residues to arginine (Radic et al., 1989; Eilat et al., 1988). The presence of arginines in the H3 loop in particular seems to confer an increased avidity for double stranded DNA. Jel 72 does not bind single stranded DNA but binds strongly to the right-handed duplex formed by poly(dG)·poly(dC). The H3 loop of Jel 72 is rich in arginines, which suggests that the mechanism of dsDNA recognition by Jel 72 may be applicable to DNA binding by the dsDNA specific autoimmune antibodies.

In order to approach the question of what are the potential interactions between an antibody and its antigen, the structures of both must be known. Fortunately, the structure of the antigen for antibodies that bind dsDNA can generally be assumed to be a helical polynucleotide that can be approximated by the 'A', 'B' or 'Z' families of DNA structures. For Jel 72, the structure of its antigen, poly(dG)•poly(dC), is assumed to be that proposed by McCall and coworkers (1985) which is derived from the crystal structure determination of d(GGGGCCCC). This structure belongs to the 'A' helix family with a deep and narrow major groove and a wide and shallow minor groove.

An examination of the electrostatic surface potential of the antibody combining region may also aid in deriving a working model for DNA binding by Jel 72. Since DNA is a polyanion the distribution of positive potential may provide clues as to the general disposition of the DNA helix with respect to the antibody CDRs. The electrostatic surface potential of the combining region of Jel 72 is displayed in Figure V-2 with contours of +/- 1 kcal/mol. The surface potential is dominated by the positive charge created by the arginine and lysine residues in CDR H3 and CDR L1. Only histidine L93 creates positive potential in CDR L3, suggesting that the sugar phosphate backbone of the DNA is interacting preferentially with the heavy chain on the CDR H3 side of the combining region.

In the crystal structures of several site specific DNA binding proteins complexed with DNA (Pavletich and Pabo, 1991; Kim et al., 1990; Luisi et al., 1991), arginines are observed to form specific hydrogen bonds with the N7 and O6 atoms of

A.

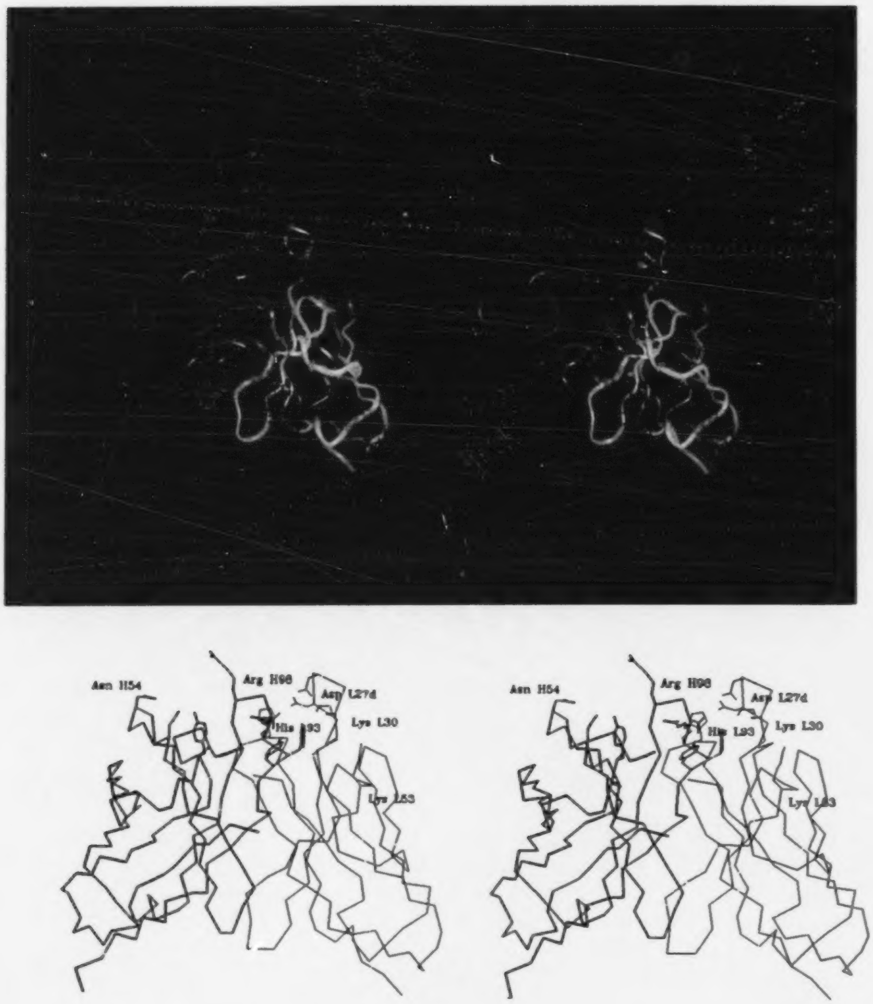

Asn H54
Arg H98
Asp L27d
His L93
Lys L30
Lys L53
Asn H54
Arg H98
Asp L27d
His L93
Lys L30
Lys L53

Figure V-2: Combining Region and Electrostatic Surface Potential of Jel 72. **A.)** View from the H3 side of the combining region. The V_H domain is shown in green (top) and in thick lines (bottom), the V_L domain in yellow (top) and in thin lines (bottom). The trace of the polypeptide backbone is shown as a solid ribbon (top) or as only $C\alpha$ atoms (bottom). Surface potentials are shown at contour levels of +1 kcal/mole (red) and -1 kcal/mole (blue), and were calculated as described in Chapter IV.

B.

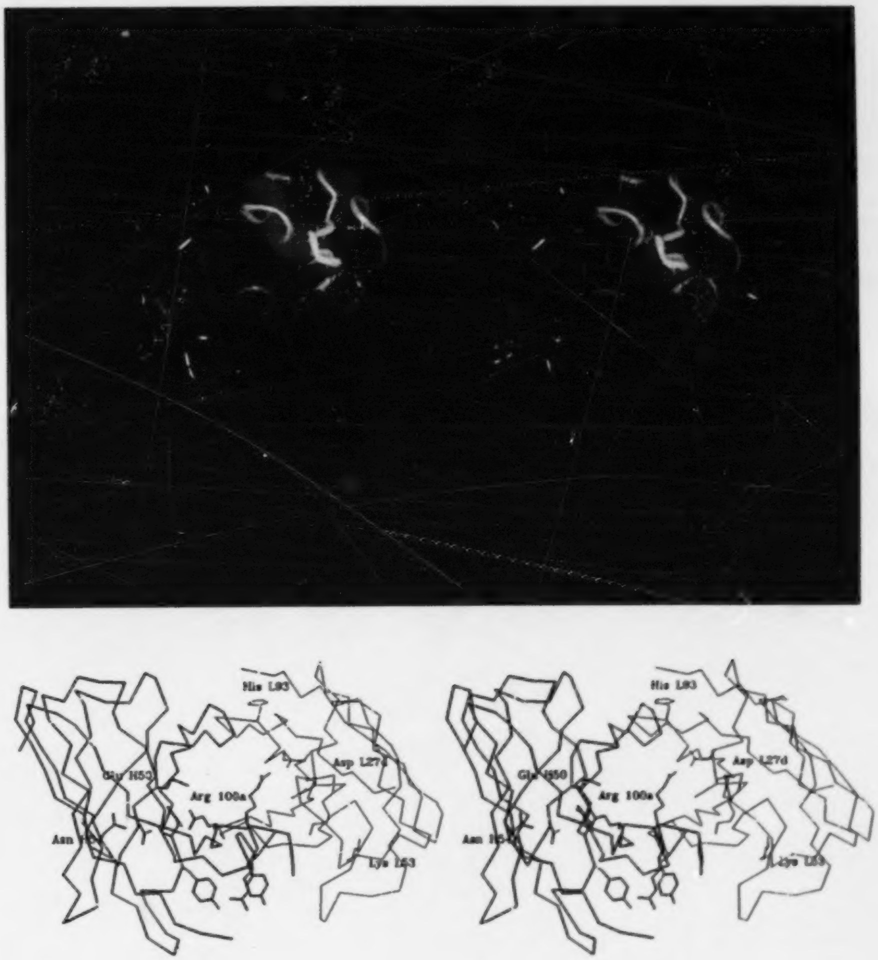


Figure V-2: Combining Region and Electrostatic Surface Potential of Jel 72. **B.)** View looking down into the combining region. The V_H domain is shown in green (top) and in thick lines (bottom), the V_L domain in yellow (top) and in thin lines (bottom). The trace of the polypeptide backbone is shown as a solid ribbon (top) or as only Cα atoms (bottom). Surface potentials are shown at contour levels of +1 kcal/mole (red) and -1 kcal/mole (blue), and were calculated as described in Chapter IV.

the guanines in the major groove of B-DNA. The arginines at the tip of the H3 loop of Jel 72 may be interacting with the guanine bases in poly(dG)·poly(dC) in a similar way, thus accounting for the extraordinary specificity for this sequence exhibited by Jel 72 (Lee et al., 1984).

Assuming that these residues in CDR H3 interact with poly(dG)·poly(dC) via the major groove, and pivoting the DNA about this point, then two potential ways in which the protein could bind to DNA are apparent. Either the long axis of the DNA helix lies approximately parallel to the V_L-V_H domain interface and spans the combining region from CDR H3 to CDR L3, or the DNA binds at an angle to the interface and interacts primarily with the V_H domain. These models for the complex between the Fv fragment and the DNA were generated manually using interactive graphics and the intermolecular "bump" option within INSIGHTII, in order to achieve an alignment of the major and minor grooves of the DNA with the antibody hypervariable loops while keeping the number of close van der Waals contacts between the two molecules at a minimum. The manually docked Fv-DNA complexes were then subjected to very limited energy minimization (approximately 15 - 20 cycles) using the energy minimization routines within X-PLOR to relieve the few bad contacts that did exist.

In the first model of the complex, the DNA spans the combining region, and the CDR H3 loop penetrates the major groove of the DNA (model 1, Figure V-3a). Residues in the CDR

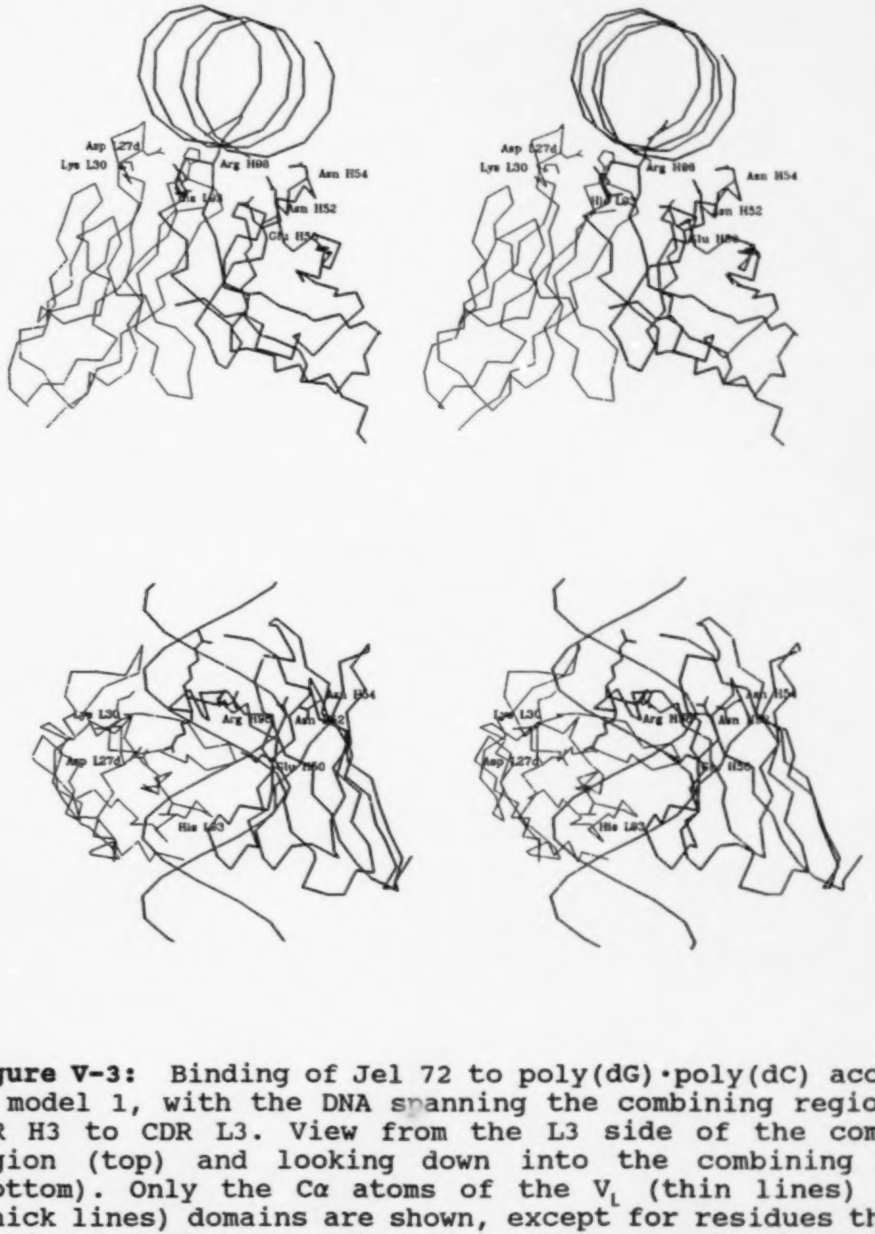

Asp L27d
Lys L30
Arg H98
Asn H54
His L83
Asn H52
Glu H50

Figure V-3: Binding of Jel 72 to poly(dG)·poly(dC) according to model 1, with the DNA spanning the combining region from CDR H3 to CDR L3. View from the L3 side of the combining region (top) and looking down into the combining region (bottom). Only the Cα atoms of the V_L (thin lines) and V_H (thick lines) domains are shown, except for residues that are mentioned in the text. The DNA is shown as only the phosphate atoms.

L3 loop (trp 89, gln 90, his 93) are in a position to make specific contacts within the minor groove of the DNA. A problem with this model arises, however, by the presence of the negative charge on glutamic acid H50 near to which the sugar phosphate backbone of the DNA would have to pass. This interaction would be highly unfavorable unless there was a cation bound to this site on the protein. The UO_2^{2+} cation binds to glutamic acid H50 with high affinity in the uranyl nitrate derivative, indicating that this may be a feasible model for DNA binding.

Site specific mutagenesis experiments can be used to test these assumptions by altering the potential hydrogen bonding residues in CDR L3 and abolishing the negative charge on glutamic acid H50 by changing it to a glutamine. These substitutions should decrease the affinity of Jel 72 for poly(dG)·poly(dC). Binding may be increased by incorporating a positive charge at position H50 by substituting glutamic acid with an arginine or by changing glutamine L92 to an arginine, thus increasing the positive potential near CDR L3.

If Jel 72 binds poly(dG)·poly(dC) at an angle (model 2, Figure V-4), then the mutations suggested above should have relatively little effect on binding. Instead, the DNA would not interact with CDR L3 and would generally avoid the negative charge of glutamic acid H50 . In this model the CDR H3 loop is again positioned within the major groove of the DNA, but it is the amino acids of the CDR H1 and CDR H2 loops

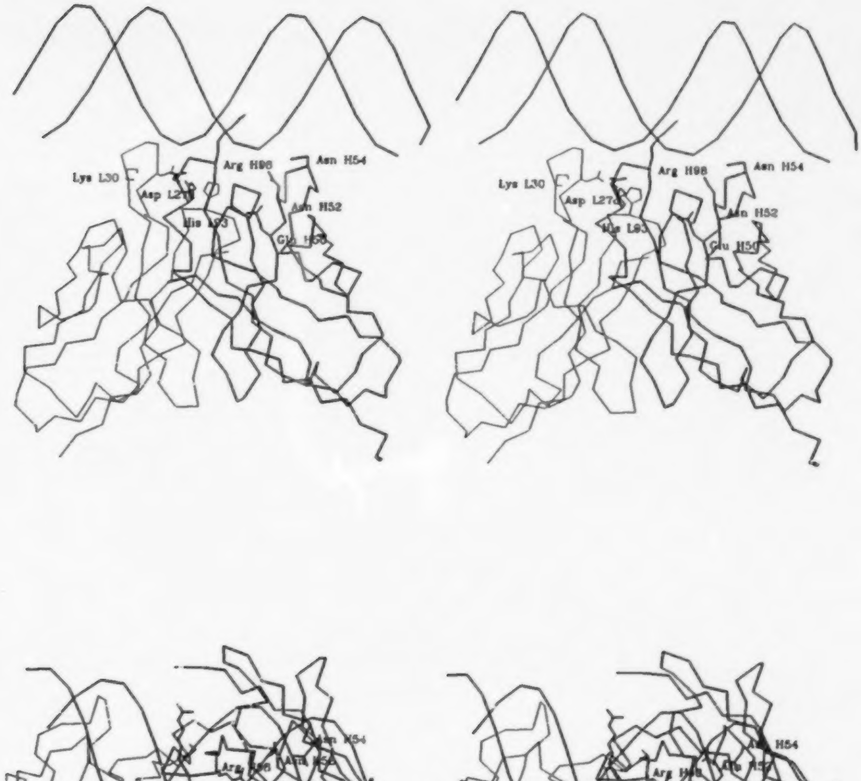

Lys L30
Asp L27d
Arg H98
Asn H54
Asn H52
His L93
Glu H50

Figure V-4: Binding of Jel 72 to poly(dG)·poly(dC) according to model 2, with the DNA inclined at an angle from CDR H3 to CDR H2. View from the L3 side of the combining region (top) and looking down into the combining region (bottom). Only the Cα atoms of the V_L (thin lines) and V_H (thick lines) domains are shown, except for residues that are mentioned in the text. The DNA is shown as only the phosphate atoms.

that are in position to interact with the minor groove and the sugar phosphate backbone of the DNA. Specific residues that may interact with the sugar phosphate backbone and bases within the minor groove include: threonines H28 and H30, serine H31 and tyrosine H32 in CDR H1, and asparagines H52 and H54, and serine H53 in CDR H2. Substituting these residues with amino acids with nonpolar side chains should greatly reduce binding affinity for dsDNA.

If Jel 72 does bind poly(dG)·poly(dC) according to model 2, then it should be possible to increase the overall affinity for DNA by increasing the positive potential of CDR H2 through the substitution of asparagine H54 with an arginine. Residues in CDR L1 that could be interacting with the sugar phosphate backbone, according to this model, include lysine L30 serine L27e and aspartic acid L28, which if replaced with a lysine or arginine should increase affinity for dsDNA.

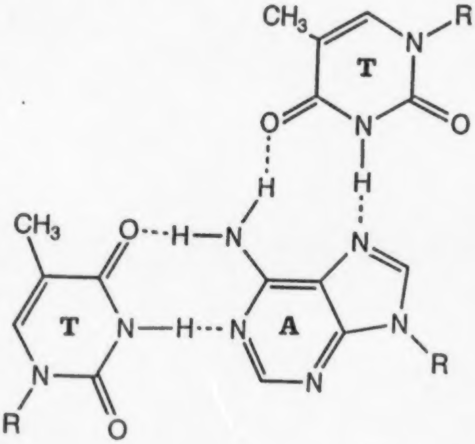
There is little to distinguish between the above two models on the basis of the surfaces and charges buried when the protein binds DNA. According to either binding model, approximately five phosphate groups on the DNA become inaccessible to a 1.7 Å radius solvent probe when bound by the Fab. These buried DNA phosphates lie mostly in the vicinity of the CDR H3 loop. With the DNA bound according to the first model, spanning the combining site (Figure V-3), a total surface area of approximately 800 $Å^2$ are buried on the Fab, while slightly more than 700 $Å^2$ of the Fab are buried by the

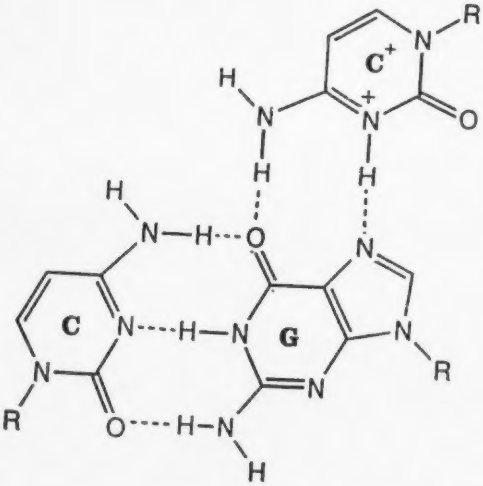
DNA bound according to the second model (Figure V-4). Although these models are by neccessity rather primitive, not taking into account any potential conformational change of the antibody CDRs upon binding poly(dG)·poly(dC), it is interesting to note that the total buried surface area of the Fab is slightly larger than the total buried surface area observed in some of the Fab-protein complexes (Davies et al., 1990; Tulip et al., 1992). Unlike most protein antigens, the DNA will extend beyond the combining region and may be interacting with areas of the antibody not normally considered to be involved in the interaction with antigen.

Regardless of the disposition of the DNA helix, both models hinge on the assumption that the residues near the apex of the CDR H3 loop bind in the major groove of the DNA. If this is the case then mutations can be made to enhance or alter the specificity of Jel 72. Replacing valine H99 with a threonine should increase the potential for forming hydrogen bonds, while replacing arginines 98 and 100A with glutamines or asparagines may increase the specificity for oligonucleotides containing adenine.

V.4 Triple Stranded DNA

Interest in triple stranded nucleic acids has grown in recent years following the discovery that it may occur in vivo and play a role in DNA replication and transcription (Htun and Dahlberg, 1988) or in the recombination and condensation of DNA in eukaryotic chromosomes (Morgan, 1979). Jel 318 recognizes triplexes of the form poly(pyr)·poly(pur)·poly(pyr) in which the two pyrimidine strands have opposite polarity. The second pyrimidine strand contains protonated cytosines that form Hoogsteen base pairs in the major groove with guanines of the purine strand. A model for the structure of this triplex is based on that determined by x-ray diffraction analysis of fibers of poly(dT)·poly(dA)·poly(dT) (Arnott and Selsing, 1974).





The underlying nature of the duplex is 'A' type, possessing a deep and narrow major groove into which the second pyrimidine strand inserts in a direction parallel to the purine strand. This presents a problem when trying to visualize antibody binding to triple stranded DNA. What remains of the major groove in the triplex is far too narrow for the antibody to make specific contacts with the edges of the DNA bases. The presence of the negatively charged phosphodiester backbone of the second pyrimidine strand in the major groove makes this an area of extremely negative electrostatic potential. It is likely, therefore, that any contacts the antibody makes with this region of the triplex will be dominated by interactions with the sugar phosphate backbone. The minor groove of the triplex, however, is wide enough to accommodate the portion of the CDR H2 loop of Jel 318 that projects outward from the combining region.

While the possibility that the antibody CDRs will undergo conformational change upon binding triplex can not be discounted, the shape of, and distribution of positive charge within, the combining region suggests a possible model for triplex recognition by Jel 318. The binding of Jel 318 to triplex has been investigated by solid phase radioimmunoassay, and these results show that the antibody has a preference for AT rich over GC rich triplex (J.S. Lee, personal communication). The binding model, presented below, proposes that this preference for AT rich triplex is achieved by the antibody making base specific contacts with the DNA via the minor groove of the DNA.

The model for the Fab-triplex complex is based on the assumption that the CDR H2 loop interacts with the minor groove of the DNA. This portion of CDR H2 possesses positive electrostatic potential, while there is another region of positive potential across the combining region centered around CDR H3 and CDR L2 (Figure V-5). The antibody would then bind the triplex at an angle with the long axis of the DNA helix approximately parallel to the V_H-V_L domain interface. The protruding CDR H2 loop can, thus, make base specific contacts with the DNA via the minor groove and the region of positive charge near CDR L2 and CDR H3, created by residues lysine L53 and arginine H98, interacts with the sugar phosphate backbone of the DNA (Figure V-6). Substituting lysine L53 with an arginine would increase the potential for forming hydrogen bonds with the DNA backbone while conserving the positive charge in this area of the combining region. This substitution should increase the overall affinity of Jel 318 for triplex DNA.

With the DNA triple helix bound in this orientation approximately six phosphate groups on the DNA become inaccessible to a 1.7 Å radius solvent probe. These DNA phosphates are located near CDR H3 in the center of the combining region. Upon complex formation a total surface area of about 700 $Å^2$ on the antibody would become buried. As with

A.

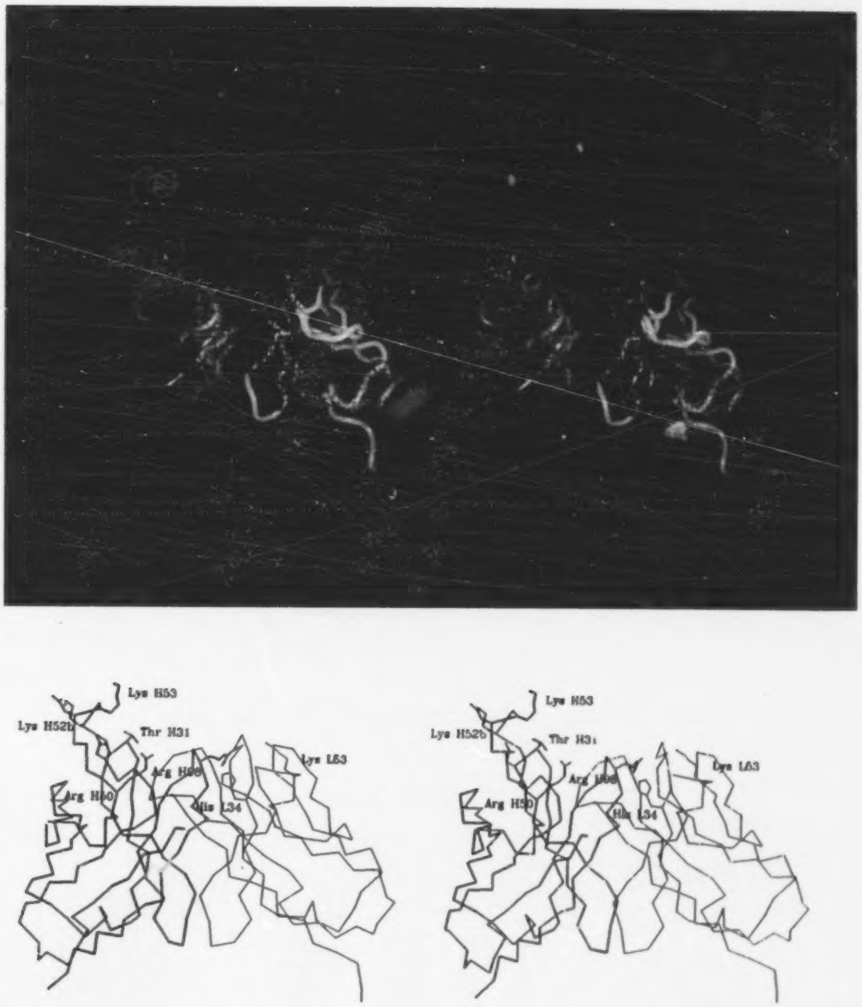

Lys H53
Lys H52b
Thr H31
Arg H98
Lys L53
Arg H50
His L34
Lys H53
Lys H52b
Thr H31
Arg H98
Lys L53
Arg H50
His L34

Figure V-5: Combining Region and Electrostatic Surface Potential of Jel 318. **A.)** View from the H3 side of the combining region. The V_H domain is shown in green (top) and in thick lines (bottom), the V_L domain in yellow (top) and in thin lines (bottom). The trace of the polypeptide backbone is shown as a solid ribbon (top) or as only Cα atoms (bottom). Surface potentials are shown at contour levels of +1 kcal/mole (red) and -1 kcal/mole (blue), and were calculated as described in Chapter IV.

B.

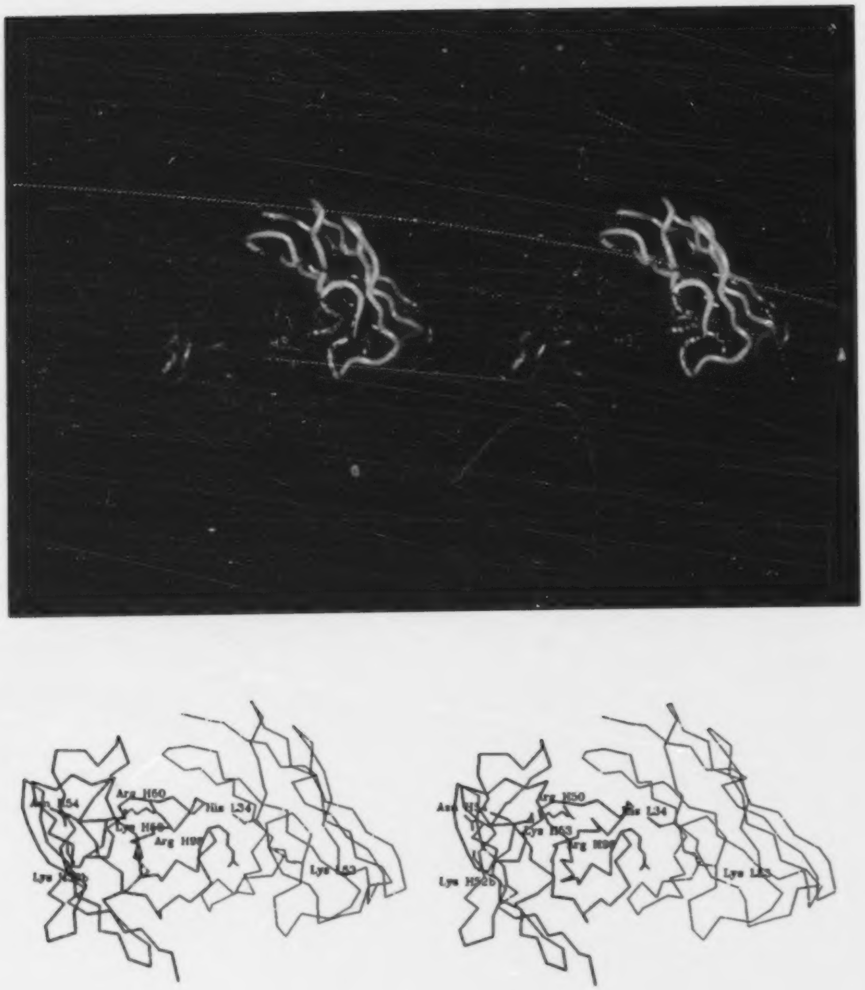


Figure V-5: Combining Region and Electrostatic Surface Potential of Jel 318. **B.)** View looking directly down onto the combining region. The V_H domain is shown in green (top) and in thick lines (bottom), the V_L domain in yellow (top) and in thin lines (bottom). The trace of the polypeptide backbone is shown as a solid ribbon (top) or as only Cα atoms (bottom). Surface potentials are shown at contour levels of +1 kcal/mole (red) and -1 kcal/mole (blue), and were calculated as described in Chapter IV.

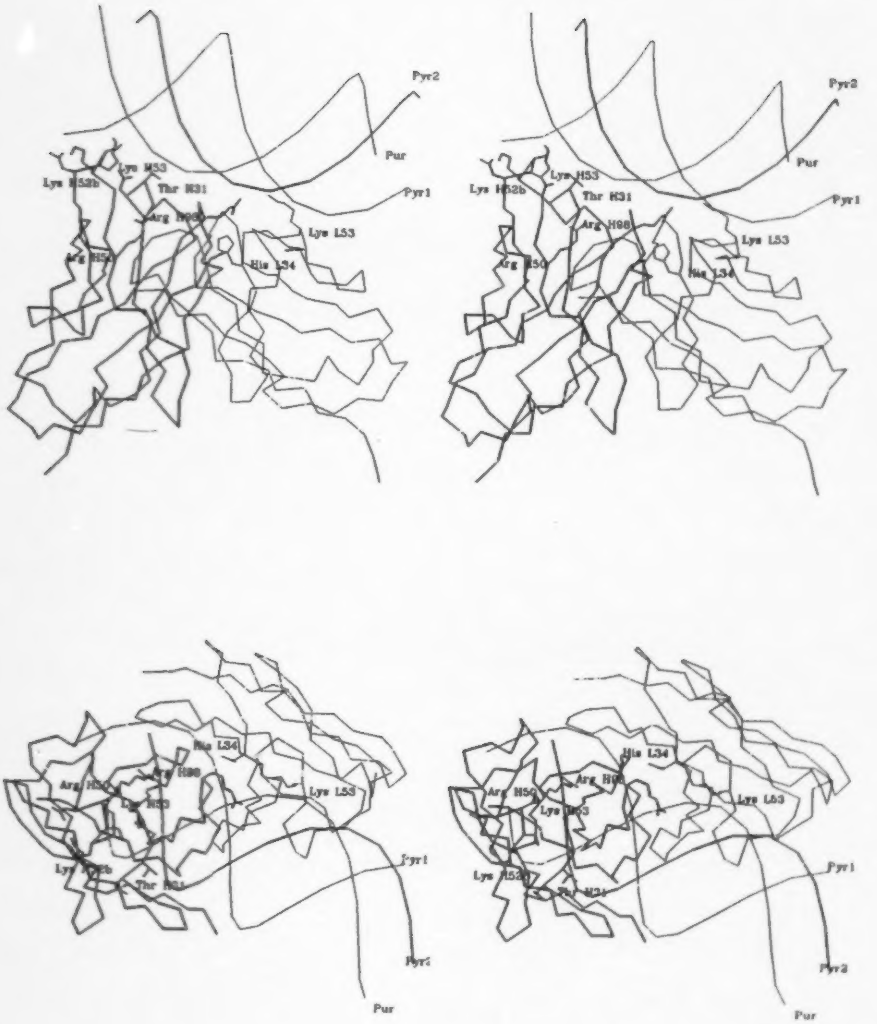

Pyr2
Pur
Pyr1
Lys H52b
Lys H53
Thr H31
Arg H98
Lys L53
Arg H50
His L34
Pyr2
Pur
Pyr1
Lys H52b
Lys H53
Thr H31
Arg H98
Lys L53
Arg H50
His L34
His L34
Arg H50
Arg H98
Lys H53
Lys L53
Pyr1
Lys H52b
Thr H31
Pyr2
Pur
His L34
Arg H50
Arg H98
Lys H53
Lys L53
Pyr1
Lys H52b
Thr H31
Pyr2
Pur

Figure V-6: Proposed model for the binding of Jel 318 to triplex DNA. View from the H3 side of the combining region (top) and looking down into the combining region (bottom). Only the Cα atoms of the V_L (thin lines) and V_H (thick lines) domains are shown, except for residues that are mentioned in the text. The DNA is shown as only the phosphate atoms.

the proposed Jel 72 DNA complexes, this amount of total buried surface area is similar to that observed in the crystal structures of Fab-protein complexes (see Davies et al, 1990).

By positioning CDR H2 in the minor groove of the triplex there is the potential for forming base specific hydrogen bonds that can explain the preference of Jel 318 for AT rich triplex. The minor groove of AT containing DNA is more acidic than GC rich triplex owing to the presence of the O2 of the thymine bases. In the Fab-triplex model residues Lys H52b, Ser H52c, Lys H53 and Asn H54 could potentially interact with atoms of the DNA in the minor groove, with the side chain of Lys H53 in position to donate a hydrogen bond to the O2 atom of thymine. In GC containing triplex the exocyclic amino group of the guanine bases hydrogen bonds with the O2 of cytosines and is incapable of accepting a hydrogen bond from the lysine side chain. The presence of the guanine NH_2 group in the minor groove may interfere with the lysine interacting with the O2 atom of cytosine. If this proposed interaction is a major determinant of sequence specificity, then Jel 318 should bind equally well to triplex containing inosine in place of guanine. Replacing Lys H53 with a residue that can accept a hydrogen bond from the NH_2 group of guanine may result in an antibody with increased affinity for GC rich triplex.

The proposed triplex binding model may explain the affinity of Jel 318 for triplex over duplex DNA. If the third strand of the DNA is not present, i.e. if the DNA is a duplex, then the residues of the L2 loop are not in a favorable position to interact with the phosphodiester backbones of the other two DNA strands. In a triplex the Hoogsteen strand could interact with these residues, if the H2 loop is penetrating the minor groove. Substituting lysine L53 with an arginine could increase the potential for forming hydrogen bonds with the DNA while conserving the positive charge in the area of CDR L2. This substitution should serve to increase the overall affinity of Jel 318 for triple stranded DNA.

There is one observation that can be made regarding the uniqueness of the combining site of Jel 318 and why other triplex specific antibodies may be difficult to obtain. Jel 318 has an unusually short L1 loop, six residues shorter than the L1 loop in both Hed10 and Jel 72. If the L1 loop of Jel 318 were longer it would interfere with the proposed interaction of the triplex with the L2 loop. It may be possible to engineer other triplex specific antibodies by simply shortening the L1 loop and substituting residues in the L2 loop with lysines or arginines.

VI. Conclusions

VI.1 Discussion

The results presented in this dissertation represent a significant advancement of our understanding of the structural basis for antibody-DNA interactions. Up until this work, our knowledge of the three dimensional structures of nucleic acid binding antibodies was restricted to the crystal structures of two single stranded DNA specific antibody Fab fragments in an uncomplexed form; Hed10 (Cygler et al., 1987), and BV04-01 (Herron et al., 1991), as well as the structure of BV04-01 in complex with the trinucleotide, d(pT_3) (Herron et al., 1991).

The crystal structure of Jel 72, however, is the first dsDNA specific antibody Fab fragment to have been determined. The structure and properties of the combining region of Jel 72 suggests possible mechanisms for its recognition of poly(dG)·poly(dC) and these mechanisms may be applicable to the binding of dsDNA by autoimmune antibodies, which are a major cause of tissue damage in the human autoimmune disorder systemic lupus erythematosus. The DNA binding models that are proposed can be tested by site specific mutagenesis experiments and the mechanisms of DNA binding by these antibodies further elucidated and used for the interpretation of biochemical and immunological results. Of the two binding models proposed, the first model, with the DNA spanning the combining region from CDR H3 to CDR L3, would seem to be the more reasonable since a larger area of the Fab becomes buried than if the DNA were to bind at an angle, as proposed for model two. The second model, however, is more energetically favourable because it places the negatively charged sugar phosphate backbone of the DNA in regions of primarily positive electrostatic potential on the surface of the Fab. This model is also similar to a model that we have proposed for the recognition of 'B' DNA by an autoimmune dsDNA binding antibody (Radic et al., 1993), in which the DNA is proposed to interact primarily with the V_H domain. Jel 72 has been cloned into Escherichia coli as a single chain Fv and experiments to test these hypotheses are currently underway (J.S. Lee, personal communication). Jel 72 shares several features with autoimmune dsDNA specific antibodies that occur in systemic lupus erythematosus, particularly the presence of arginines in the CDR H3 hypervariable loop. Insights gained into antibody-DNA recognition by these studies may be pertinent to the diagnosis and treatment of this disorder.

The three dimensional structure of Jel 318 is the first of a triple-stranded DNA binding antibody Fab fragment, or of any triplex specific protein, to have been determined. The combining region presents a relatively flat surface characterized by a single protruding loop that extends away from the main body of the molecule. A model for the Fab-triplex complex is proposed that places this protruding loop in the minor groove of the DNA, where the residues in CDR H2 of the antibody may be able to make base specific contacts with the triplex. Site specific mutagenesis experiments to test this model are proposed and will be facilitated by the fact that Jel 318 has been cloned as a single chain F_v into Escherichia coli. These experiments will help to further delineate the mechanism of triplex binding by Jel 318 and may aid in the design of triplex specific antibodies with altered or enhanced specificity for different triple-stranded nucleic acids.

When we began this project we were enticed by the many apparent advantages of studying protein-nucleic acid interactions within the framework of the antibody molecule. The first, and foremost, advantage is the ability to obtain homogenous antibody preparations in large quantities. This was a necessary requirement because over the years we have consumed vast quantities of antibody preparations for use in the production of Fab fragments for the crystallization trials and co-crystallization experiments with various with oligonucleotides.

Another advantage of using antibodies to study protein-nucleic acid interactions is that diffraction quality crystals of antibody Fab fragments are relatively easy to obtain. Although the digestion of antibodies with papain produces an isoelectrically heterogenous mixture of Fab fragments, pure preparations of Fab fragments can be obtained in a relatively straightforward manner, as the results presented in Chapter II illustrate. However, purification of the nearly identical

Fab molecules was not simple and required rather drastic conditions, such as the need to load Jel 72 Fab fragments onto the column at pH 10.0. Even with the use of extremely shallow NaCl gradients to elute the Fabs from the column, the purification protocols never succeeded in achieving very clean separations of the Fab peaks. This resulted in a need to repetitively purify the Fab fragments and significantly lowered the yields of pure material available for use in crystallization trials.

An unexpected aspect of the production of Fab fragments was the unreliability of the papain preparations. We invariably used only the freshest papain in the preparative digestions. Older papain preparations could not be relied upon to give reproducible results. Also, the purified Fab fragments were apparently unstable. Once purified the Fabs had to be used almost immediately in crystallization experiments. Older preparations tended to degrade into mixtures of Fab species that no longer gave suitable crystals.

Nevertheless, isoelectrically pure Fab species could be obtained in a single purification step. As illustrated in Figure II-3 for Jel 72, the use of purified Fab species in the crystallization experiments had a dramatic effect on the size and quality of the crystals obtained. Once the purification conditions have been determined isoelectrically pure Fabs can be routinely obtained. We suggest that any crystallization experiments performed with antibody Fab fragments, whether alone or in complex with low molecular weight haptens or larger biological antigens, be preceded by an examination of the isoelectric purity of the protein preparations. For that matter, crystallization experiments with any biological macromolecule should be preceded by a careful examination of the isoelectric purity of the sample.

Another major advantage in using antibodies to study protein-nucleic acid interactions is that the structures of the Fab fragments can be solved easily by molecular replacement. Although we eventually solved all of the structures using these techniques, the interpretation of the rotation and translation function results was not without complications. The known variation in the elbow angle relating the variable and constant domain pairs led us to employ a protocol involving separating search models into the individual domain pairs for use in the molecular replacement calculations. Invariably the constant domain pairs were much easier to place in the cell than the variable domain pairs. The interpretation of the results was facilitated by orienting the model Fab fragments in such a way that the results obtained for the constant domain pairs could be used to decipher the correct solution for the variable domain pairs. Further complications arose by having two Fabs in the asymmetric unit for Jel 72 in crystal form I, and by the nearly isomorphous crystals of Jel 318 and Fab Kol despite their different packing arrangements.

One impediment we encountered to the rapid solution and refinement of the crystal structures was the lack of amino acid sequence information for the variable domains. The resolution of the data was not high enough and the structures not well-behaved, showing considerable disorder in the solvent exposed loop regions, that the interpretation of the electron density in the absence of sequence information was fraught with errors. The initial models for the H3 and L1 loops of Jel 72 were several residues shorter than the actual loops owing to the poor quality of their electron density. For Jel 318 the electron density was poor for long stretches of exposed loops in the constant domains. Without having a good model in place for these domains the density for the variable domains was also poor. Once the sequence for the variable domains became known, these residues could be fit to the density and the clarity of the subsequent maps for the entire molecule improved substantially. To circumvent this problem, we suggest that any crystal structure determination of an antibody Fab fragment be preceded by the determination of the amino acid sequence of the variable domains.

An exciting development that we did not expect at the beginning of the project was the possibility of modelling antibody combining region based on their amino acid sequences. When we received the sequence information we noted that nearly all of the hypervariable regions belonged to one of the canonical structure classes for immunoglobulin loops as described by Chothia and Lesk (1987). In order to avoid the appearance of bias we modelled the structures of Jel 72 and Jel 318 before proceeding with their refinements, and applied the modelling techniques in as rational and unbiased a manner as possible. The results presented in Chapter IV seem to indicate that good working models for antibody combining regions can generally be obtained by applying a relatively simple modeling protocol. Considering that a model for an antibody combining region can be derived in less than one day, and that a crystal structure determination can take several months at best, perhaps a modelled combining site can serve to clarify biochemical and immunological data and help to elucidate antibody antigen interactions at a considerable savings in both time and manpower.

We have, in fact, applied the modelling protocol to over a dozen other DNA binding antibodies. In collaboration with the laboratory of Martin Weigert at the Fox Chase Cancer Center we have modelled the combining regions of murine autoimmune antibodies isolated from mice suffering from a disorder similar to SLE in humans. Unlike the antibodies that we have characterized crystallographically, these antibodies have affinities for both single stranded and double stranded DNA (Shlomchik et al., 1987). These antibodies were found to use related V_H genes and the propensity for binding double stranded DNA was achieved by somatic mutation of hypervariable residues. Replacement of residues with arginine, particularly in the H3 loop, correlates with increased binding to double stranded DNA. Some of the antibodies were highly specific for single stranded DNA, while the antibodies that bound double stranded DNA retained their affinity for single stranded DNA. None of the antibodies bound double stranded DNA as well as either Jel 72, Jel 274 or Jel 242.

These results suggested that the binding of DNA by these antibodies was dominated by non-specific interactions with the sugar phosphate backbone. The models for the combining regions of the antibodies that were single strand specific, and showed no affinity for double stranded DNA, were similar in many respects to the combining region of Hed10. Their combining sites are characterized by a long cleft roughly parallel to the $V_H \cdot V_L$ domain interface. The interior of this cleft is usually enriched in aromatic amino acids while the ridges on either side of the cleft often contain basic amino acids. For these antibodies models for their interaction with DNA can be derived that are generally similar to that observed in the BV04-01 complex with $d(pT_3)$ (Herron et al., 1991) and to the proposed interactions of Hed10 with poly(dT). The DNA bases can stack with the aromatic amino acid side chains within the cleft while the negatively charged phosphate groups interact with the regions of positive electrostatic potential. The flexible nature of single stranded DNA and the possibility of antigen induced conformational change of the hypervariable loops makes it unlikely that more elegant and detailed models for DNA binding by these autoimmune antibodies can be derived. The similarities noted between the combining regions of Hed10 and BV04-01, however, suggest that there may be some generalizations that can be made for antibodies that bind single stranded nucleic acids. The recurrence of a histidine in the L1 loop, observed to hydrogen bond with a sugar of the DNA, and an arginine in the H2 loop, interacting with the DNA phosphate groups, may serve to orient the nucleic acid in such a way that the bases are in a position to make contacts with residues of the H3 loop.

The models for the combining regions of the autoimmune antibodies that also bind to double stranded DNA are more similar to Jel 72 than to Hed10. Since an increased affinity of these antibodies for double stranded DNA is correlated with somatic mutation of hypervariable residues to arginine, it is not surprising that their modelled combining sites possess much more positive electrostatic potential. The likely manner in which these antibodies bind DNA is in a more or less parallel fashion as proposed for model one of Jel 72 binding poly(dG)·poly(dC). However, we cannot exclude the possibility that at least some of these antibodies are binding to double stranded DNA in a perpendicular fashion more similar to the second model proposed for Jel 72 and poly(dG)·poly(dC). In either case, antibody contacts to the DNA are probably dominated by non-specific interactions with the sugar phosphate backbone. It is possible, even probable, that within the vast repertoire of double stranded DNA specific antibodies in nature both of these models for binding are employed. The crystallographic studies of antibody-antigen and protein-DNA complexes have established that surface complementarity between the two molecules is an important aspect of their interactions. There is no reason to expect that antibody-DNA complexes should not also have reasonably complementary surfaces. Even if Jel 72 does not bind to DNA as proposed in the second model, it should be possible to engineer antibodies that do. It may even be possible to obtain antibodies that bind to specific DNA sequences, similar to the site specific DNA binding proteins that regulate genetic transcription, rather than antibodies that recognize repeating DNA polymers.

The fact that all three antibody Fab fragments that we have examined crystallographically bind to repeating DNA sequences has frustrated attempts at obtaining co-crystals of Fab fragments bound to DNA. The antibodies do not bind the small oligonucleotides used in the co-crystallization experiments particularly well and there is doubt that the antibody is recognizing a single site of the repeating DNA sequence. In, for example, a twelve nucleotide long piece of DNA the antibody may be binding the central six base pairs or to six base pair long stretches anywhere along the DNA. The end result is that the concentration in solution of a single antibody-DNA complex is very low. The equilibrium between free Fab and DNA and Fab-DNA complex is shifted far enough towards the uncomplexed molecules that the most likely crystals one will obtain are of the Fab alone. Such was the case for the second crystal form of Jel 72. If as the result of site specific mutagenesis experiments, as suggested in Chapter V, antibodies can be engineered with a higher affinity for unique DNA sequences, then perhaps these antibodies will prove to be more amenable for study in co-crystallization experiments with DNA.

Nevertheless, we have derived what we believe to be reasonable models for DNA binding by these antibodies. These models can be easily tested and refined as more biochemical and immunological data become available concerning the roles specific residues of the antibody are playing in the recognition of nucleic acid antigens. The apparent success of the modelling of antibody combining regions would lead us to suggest that as the amino acid sequences of other DNA binding antibodies are determined models for their combining regions should also be derived. By a comparison of these model structures with the structures of the Fab fragments we have determined by x-ray crystallography important insights may be gained into antibody nucleic acid interactions.

VII. Bibliography

D. Abu, J. Dobson, and D.G. Williams (1981) Clin. Exp. Immunol. **43**:605.

A.K. Aggarwal, D.W. Rogers, M. Drotter, M. Ptashne, and S.C. Harrison (1988) Science **242**:899.

A.G. Amit, R.A. Mariuzza, S.E.V. Phillips, and R.J. Poljak (1986) Science **233**:747.

L.M. Amzel and R.J. Poljak (1979) Annu. Rev. Biochem. **48**:961.

W.F. Anderson, D.H. Ohlendorf, Y. Takeda, and B.W. Matthews (1981) Nature **290**:754.

W.F. Anderson, M. Cygler, R.P. Braun, and J.S. Lee (1988) Bioessays **8**:69.

S. Arnott and E. Selsing (1974) J. Mol. Biol. **88**:509.

S. Arnott, P.J. Bond, E. Selsing, and P.J.C. Smith (1976) Nucl. Acids Res. **11**:4141.

P.J. Artymiuk, C.C.F. Blake, D.W. Rice, and K.S. Wilson (1982) Acta Cryst. **B38**:778.

D.W. Ballard and E.W. Voss Jr. (1985) J. Immunol. **135**:3372.

S.M. Behar and M.D. Scharff (1988) Proc. Natl. Acad. Sci. USA **85**:3970.

F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rogers, O. Kennard, T. Shimanouchi, and M. Tasumi (1977) J. Mol. Biol. **112**:535.

M.M. Berry, C.D. Mol, W.F. Anderson, and J.S. Lee (1993) J. Biol. Chem., submitted for publication.

T.N. Bhat (1988) J. Appl. Cryst. **21**:279.

T.N. Bhat, G.A. Bentley, T.O. Fischman, G. Boulot, and R.J. Poljak (1990) Nature **347**:483.

F. Birg, D. Praseuth, A. Zerial, N.T. Thuong, U. Asseline, T. Le Doan, and C. Helene (1990) Nucl. Acids Res. **18**:2901.

H.C. Birnboim, R.R. Sederoff, and M.C. Patterson (1979) Eur. J. Biochem. **98**:301.

T.L. Blundell and L.N. Johnson (1976) In Protein Crystallography, Academic Press, New York.

A. Boodhoo, C.D. Mol, J.S. Lee, and W.F. Anderson (1988) J. Biol. Chem. **263**:18578.

S. Bram (1971) Nature New Biol. **233**:161.

R.P. Braun and J.S. Lee (1986) Nucl. Acids. Res. **14**:5049.

R.P. Braun and J.S. Lee (1987) J. Immunol. **139**:175.

R.P. Braun amd J.S. Lee (1988) J. Immunol. **141**:2084.

J.N. Breg, J.H.J. van Opheusden, M.J.M. Burgering, R. Boelens, and R. Kaptein (1990) Nature **349**: 586.

M.M. Brigido and B.D. Stollar (1991) J. Immunol. **146**:2005.

R.E. Bruccoleri and M. Karplus (1987) Bioploymers **26**:132.

A. Brünger, K. Kuriyan, and M. Karplus (1987) Science **235**:458.

A. Brünger, M. Karplus, and G.A. Petsko (1989) Acta Cryst. **A45**:50.

G.D. Burkholder, L.J.P. Latimer, and J.S. Lee (1991) Chromosoma **101**:11.

M. Cooney, G. Czernuszewicz, E.H. Postal, S.J. Flint, and M.E. Hogan (1988) Science **241**:456.

C. Chothia and J. Janin (1981) Proc. Natl. Acad. Sci. USA **78**:4146.

C. Chothia and J. Janin (1982) Biochemistry **21**:3955.

C. Chothia, A.M. Lesk, M. Levitt, A.G. Amit, R.A. Mariuzza, S.E.V. Phillips, and R.J. Poljak (1986) Science **233**:755.

C. Chothia, J. Novotny, R. Bruccoleri, and M. Karplus (1985) J. Mol. Biol. **186**:651

C. Chothia and A.M. Lesk (1987) J. Mol. Biol. **196**:901.

C. Chothia, D.R. Boswell, and A.M. Lesk (1988) EMBO. J. **7**:3745.

C. Chothia, A.M. Lesk, A. Tramantano, M. Levitt, S.J. Smith-Gill, G. Air, S. Sheriff, E.A. Padlan, D.R. Davies, W.R. Tulip, P.M. Colman, S. Spinelli, P.M. Alzari, and R.J. Poljak (1989) Nature **342**:877.

P.M. Colman, W.G. Laver, J.N. Varghese, A.T. Baker, P.A. Tulloch, G.M. Air, and R.G. Webster (1987) Nature **326**:358.

P.M. Colman (1988) Adv. Immunol. **43**:99.

P.M. Colman, W.R. Tulip, J.N. Varghese, P.A. Tulloch, A.T. Baker, W.G. Laver, G.M. Air, and R.G. Webster (1989) Phil. Trans. R. Soc. London **B323**:511.

M.L. Connolly (1983) Science **221**:709.

R.A. Crowther (1972) In Molecular Replacement Method (Rossman, M.G., ed.) pp. 173-178, Gordon and Breach, New York.

B.A. Cunningham, J.S. Hemperley, B.A. Murray, E.A. Prediger, R. Brackenbury, and G.M. Edelman (1987) Science **236**:799.

M. Cygler, A. Boodhoo, J.S. Lee, and W.F. Anderson (1987) J. Biol. Chem. **262**:643.

M. Cygler and W.F. Anderson (1988) Acta Cryst. **A44**:38.

M. Cygler and M. Desrochers (1989) Acta Cryst **A45**:563.

M. Cygler, D.R. Rose, and D.R. Bundle (1992) **253**:442.

D.R. Davies and D.M. Segal (1971) Met. Enz. **22**:266.

D.R. Davies, S. Sheriff, and E.A. Padlan (1988) J. Biol. Chem. **263**:10541.

D.R. Davies, E.A. Padlan, and S. Sheriff (1990) Annu. Rev. Biochem. **59**:439.

M.M. Davis, K. Calame, P.W. Early, D.L. Divant, R. Joho, I.L. Weisman, and L. Hood (1980) Nature **283**:733.

M.M. Davis and P.J. Bjorkman (1988) Nature **334**:395.

J. Deisenhofer (1981) Biochemistry **20**:2361.

P. de la Paz, B.J. Sutton, M.J. Darsley, and A.R. Rees (1986) EMBO J. **5**:415.

Z.S. Derewenda, U. Derewenda, and Guy G. Dodson (1992) J. Mol. Biol. **227**:818.

R.E. Dickerson (1987) In Unusual DNA Structures, (R.D. Wells and S.C. Harvey, eds.), pp. 287-306, Springer-Verlag, New York.

R. Dziarski (1982) J. Immunol. **128**:1026.

P. Early, H. Huang, M. Davis, K. Calame, and L. Hood (1980) Cell **19**:981.

D. Eilat, B.M. Webster, and A.R. Rees (1988) J. Immunol. **141**:1745.

D. Eilat and R. Fischel (1991) J. Immunol. **147**:261.

S.C.R. Elgin (1982) Nature **300**:402.

G. Felsenfeld, D.R. Davies, and A. Rich (1957) J. Am. Chem. Soc. **79**:2023.

G. Felsenfeld and H.T. Miles (1967) Annu. Rev. Biochem. **36**:407.

R.M. Fine, H. Wang, P.S. Shenkin, D.L. Yarmush, and C. Levinthal (1986) Proteins: Struct. Funct. and Genet. **1**:342.

G.J. Fournie, P.H. Lambert, and P.A. Miescher (1974) J. Exp. Med. **140**:1189.

M. Fujinaga and R. Read (1987) J. Appl. Cryst. **20**:517.

K.C. Garcia, P.M. Ronco, P.J. Verroust, A.T. Brünger, and L.M. Amzel (1992) Science **257**:502.

E.D. Getzoff, J.A. Tainer, R.A. Lerner, and H.M. Geysen (1988) Adv. Immunol. **43**:1.

J.N.M. Glover and D.E. Pulleyblank (1990) J. Mol. Biol. **215**:653.

D.W. Green, N.M. Ingram, and M.F. Perutz (1954) Proc. Roy. Soc. **A225**:287.

K.J. Hampel, P. Crosson, and J.S. Lee (1991) Biochemistry **30**:4455.

D. Harker (1956) Acta Cryst. **9**:1.

E.N. Harris, A.E. Gharavi, M.L. Boey, B.M. Patel, C.G. Mackworth-Young, S. Loizou, and G.R.V. Hughes (1983) Lancet **2**:1211.

W.A. Hendrickson and J.H. Konnert (1981) In Biomolecular Structure, Conformation, Function and Evolution (Srinivasan, R., ed.), vol. 1, pp. 43-57, Pergamon, Oxford.

J.N. Herron, X.M. He, A.L. Gibson, E.W. Voss Jr., and A.B. Edmundsen (1987) Fed. Proc. **46**:2204.

J.N. Herron, X.M. He, M.L. Mason, E.W. Voss Jr., and A.B. Edmundsen (1989) Proteins: Struct. Funct. and Genet. **5**:271.

J.N. Herron, X.M. He, D.W. Ballard, P.R. Blier, P.E. Pace, A.L.M. Bothwell, E.W. Voss Jr., and A.B. Edmundsen (1991) Proteins: Struct. Funct. and Genet. **11**:159.

T. Honjo (1983) Annu. Rev. Immunol. **1**:499.

L. Hood, M. Kronenberg, and T. Hunkapiller (1985) Cell **40**:225.

F.B. Howard, J. Frazier, M.N. Lipsett, and H.T. Miles (1964) Biochem. Biophys. Res. Comm. **17**:93.

A. Howard, C. Neilsen, and Ng.H. Xuong (1985) In Methods in Enzymology vil 114: Part A, pp. 452-472.

H. Htun and J.E. Dahlberg (1988) Science **241**:1791.

A. Jack and M. Levitt (1977) Acta Cryst. **A34**:931.

B.H. Johnston (1988) Science **241**:1800.

S.R. Jordan and C.O. Pabo (1988) Science **242**:893.

E.A. Kabat, T.T. Wu, M. Reid-Miller, H.M. Perry, and K.S. Gottesman (1987) In Sequences of Proteins of Immunological Interest, 4th ed., United States Department of Health and Human Services, National Institutes of Health, Bethesda, Maryland.

M. Karplus and J.A. MaCammon (1983) Annu. Rev. Biochem. **52**:263.

Y. Kim, J.C. Grable, R. Love, P.J. Greene, and J.M. Rosenberg (1990) Science **249**:1307.

S. Kirkpatrick, C.D. Jelatt Jr., and M.P. Vecchi (1983) Science **220**:671.

C.R. Kissinger, B. Liu, E. Martin-Blanco, T.B. Kornberg, and C.O. Pabo (1990) Cell **63**:579.

K.L. Knight, J.U. Bowie, A.K. Vershon, R.D. Kelley, and R.T. Sauer (1990) J. Biol. Chem. **264**:3638.

G. Koehler and C. Milstein (1975) Nature **256**:495.

T. Koike, H. Tamioka, and A. Kumagai (1982) Clin. Exp. Immunol. **50**:298.

J.H. Konnert and W.A. Hendrickson (1980) Acta Cryst. **A36**:344.

J. Kuriyan, A.T. Brünger, M. Karplus, and W.A. Hendrickson (1989) Acta Cryst. **A45**:396.

E.M. Lafer, A. Moller, A. Nordheim, B.D. Stollar, and A. Rich (1981) Proc Natl. Acad. Sci. USA **78**:3546.

E.M. Lafer and B.D. Stollar (1984) J. Biomol. Struct. and Dyn. **2**:487.

A. Lange (1978) Clin. Exp. Immunol. **31**:472.

A. Larsen and H. Weintraub (1982) Cell **29**:609.

M.B. Lascombe, P.M. Alzari, G. Boulot, P. Saludjian, P. Tougard, C. Berek, S. Haba, E.M. Rosen, A. Nisonoff, and R.J. Poljak (1989) Proc. Natl. Acad. Sci. USA **86**:607.

B. Lee and F.M. Richards (1971) J. Mol. Biol. **55**:379.

J.S. Lee, D.A. Johnson, and A.R. Morgan (1979) Nucl. Acids Res. **6**:3073.

J.S. Lee, D.F. Dombroski, and T.R. Mosmann (1982) Biochemistry **21**:4940.

J.S. Lee, M.L. Woodsworth, and L.J.P. Latimer (1984) Biochemistry **23**:3277.

J.S. Lee, G.D. Burkholder, L.J.P. Latimer, B.L. Haug, and R.P. Braun (1987) Nucl. Acids Res. **15**:1047.

M. Lipsett (1964) J. Biol. Chem. **239**:1256.

B.F. Luisi, W.X. Xu, Z. Otwinowski, L.P. Freedman, K.R. Yamamoto, and P.B. Sigler (1991) Nature **352**:497.

V. Luzzati (1952) Acta Cryst. **5**:802.

V.I. Lyamichev, S.M. Mirkin, and M.D. Frank-Kamenetskii (1986) J. Biomol. Struct. Dyn. **3**:667.

L.J. Maher III, B. Wold, and P.B. Dervan (1990) Science **245**:725.

M. Marquart, J. Deisenhofer, R. Huber, and W. Palm (1980) J. Mol. Biol. **141**:369.

A.C.R. Martin, J.C. Cheetham, and A.R. Rees (1989) Proc. Natl. Acad. Sci. USA **86**:9268.

M. Matsushima, M. Marquart, T.A. Jones, P.M. Colman, K. Bartels, R. Huber, and W. Palm (1977) J. Mol. Biol. **121**:441.

B.W. Matthews (1968) J. Mol. Biol. **33**:491.

M. McCall, T. Brown, and O. Kennard (1985) J. Mol. Biol. **183**:385.

J.A. McClarin, C.A. Frederick, B.C. Wang, P.J. Greene, and J.M. Rosenberg (1986) Science **234**:1526.

D.B. McKay and T.A. Steitz (1981) Nature **290**:744.

A. McPherson (1976) Methods of Biochemical Analysis **290**:744.

H. Metzger and N. Otchin (1971) In Progress in Immunology, (B. Amos, ed.), pp. 253-267, Academic Press, New York.

J. Miller, A.D. McLachlin, and A. Klug (1985) EMBO J. **4**:1609.

S.M. Mirkin, V.I. Lyamichev, K.N. Drushlyak, V.N. Dobrynin, S.A. Filippov, and M.D. Frank-Kamenetskii (1987) Nature **330**:495.

C.D. Mol, M. Cygler, A.K.S. Muir, J.S. Lee, and W.F. Anderson (1993) J. Biol. Chem., submitted for publication.

C.D. Mol, A.K.S. Muir, J.S. Lee, and W.F. Anderson (1993) J. Biol. Chem., submitted for publication.

A. Mondragon and S.C. Harrison (1991) J. Mol. Biol. **219**:321.

M.M.W. Mooren, D.E. Pulleyblank, S.S. Wijmenga, M.J.J. Blommers, and C.W. Hilbers (1990) Nucl. Acids Res. **18**:6523.

A.R. Morgan (1979) Trends in Bioch. Sci. **10**:244

A.R. Morgan and R.D. Wells (1968) J. Mol. Biol. **37**:63.

J. Moult and M.N.G. James (1986) Proteins: Struct. Funct. and Genet. **1**:146.

T.W. Munns and M.K. Liszewski (1980) Prog. Nucl. Acids Res. Mol. Biol. **24**:109.

T.W. Munns, M.K. Liszewski, anf B.H. Hahn (1984) Biochemistry **23**:2964

M.A. Navia, D.M. Segal, E.A. Padlan, D.R. Davies, N. Rao, S. Rudikoff, and M. Potter (1979) Proc. Natl. Acad. Sci. USA **76**:4071.

M.L. Ng, J.S. Pederson, B.H. Toh, and E.G. Westaway (1983) Arch. Virology **79**:177.

E.A. Padlan, D.M. Segal, T.F. Spande, D.R. Davies, S. Rudikoff, and M. Potter (1973) Nature New Biol. **145**:165.

E.A. Padlan, E.W. Silverton, S. Sheriff, G.H. Cohen, S.J. Smith-Gill, and D.R. Davies (1989) Proc. Natl. Acad. Sci. USA **86**:5938.

E.A. Padlan (1990) Proteins: Struct. Funct. and Genet. **7**:112.

A.L. Patterson (1934) Phys. Rev. **46**:372.

N.P. Pavletich and C.O. Pabo (1991) Science **252**:809.

S.E.V. Phillips (1991) Current Opinions in Structural Biology **1**:89.

D.S. Pilch, C. Levenson, and R.H. Shafer (1990) Proc. Natl. Acad. Sci. USA **87**:1942.

M.C. Poirer (1981) J. Natl. Cancer Inst. **67**:515.

M.J.D. Powell (1977) Mathematical Programming **12**:240.

T.J. Povsic and P.B. Dervan (1990) J. Amer. Chem. Soc. **112**:9428.

D. Praseuth, L. Perroualt, T. Le Doan, M. Chassignol, N. Thuong, and C. Helene (1988) Proc. Natl. Acad. Sci. USA **85**:1349.

M.Z. Radic, M.A. Mascelli, J. Erikson, H. Shan, M. Shlomchik, and M. Weigert (1989) Cold Spring Harbor Symp. Quant. Biol. **54**:933.

M.Z. Radic, M.A. Mascelli, J. Erikson, H. Shan, and M. Weigert (1991) J. Immunol. **146**:176.

M.Z. Radic, J. Mackle, J. Erikson, C. Mol, W.F. Anderson, and M. Weigert (1993) J. Immunol, submitted for publication.

J.B. Rafferty, W.S. Sommers, I. Saint-Girons, and S.E.V. Phillips (1989) Nature **341**:705.

P. Rajagopal and J. Feigon (1989) Biochemistry **28**:7282.

G.N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan (1963) J. Mol. Biol. **7**:95.

R. Read (1986) Acta Cryst. **A42**:140.

R.J. Read and A.J. Schierbeek (1988) J. Appl. Cryst. **21**:490.

M.G. Rossman and D.M. Blow (1962) Acta Cryst. **15**:24.

M.G. Rossman (1990) Acta Cryst. **A46**:73.

S. Rudikoff and J.G. Pumphrey (1986) Proc. Natl. Acad. Sci. USA **83**:7875.

W. Saenger (1984) In Principles of Nucleic Acid Structure, New York, Springer-Verlag, p. 77.

H. Sano and C. Morimota (1982) J. Immunol. **126**:538.

Y. Satow, G.H. Cohen, E.A. Padlan, and D.R. Davies (1986) J. Mol. Biol. **190**:593.

F.A. Saul and R.J. Poljak (1992) Proteins: Struct. Funct. and Genet. **14**:363.

E. Schon, T. Evans, J. Welsh, and A. Efstratiadis (1983) Cell **35**:837.

N.C. Seeman, J.M. Rosenberg, and A. Rich (1976) Proc. Natl. Acad. Sci. USA **73**:804.

D.M. Segal, E.A. Padlan, G.H. Cohen, S. Rudikoff, M. Potter, and D.R. Davies (1974) Proc. Natl. Acad. Sci. USA **71**:4298.

S. Sheriff, E.W. Silverton, E.A. Padlan, G.H. Cohen, S.J. Smith-Gill, B.C. Finzel, and D.R. Davies (1987) Proc. Natl. Acad. Sci. USA **84**:8075.

M.J. Shlomchik, A.H. Aucoin, D.S. Pisetsky, and M.G. Weigert (1987) Proc. Natl. Acad. Sci. USA **84**:9150.

M.J. Shlomchik, M. Mascelli, H. Shan, M.Z. Radic, D. Pisetsky, A. Marshak-Rothstein, and M. Weigert (1990) J. Exp. Med. **171**:265.

J.T. Sibley, R.P. Braun, and J.S. Lee (1986) Clin. Exp. Immunol. **64**:563.

J.T. Sibley, L.J.P. Latimer, and J.S. Lee (1988) J. Immunol. **140**:3502.

S.J. Smith-Gill, C. Mainhart, T.B. Lavoie, R.J. Feldman, W. Drohan, and B.R. Brooks (1987) J. Mol. Biol. **194**:713.

R.L. Stanfield, T.M. Fieser, R.A. Lerner, and I.A. Wilson (1990) Science **248**:712.

T.A. Steitz (1990) Quart. Rev. Biophys. **23**:205.

B. Stiepe, A. Plückthun, and R. Huber (1992) J. Mol. Biol. **225**:739.

B.D. Stollar and V. Stollar (1970) Virology **42**:276.

B.D. Stollar (1987) CRC Crit Rev. Biochem. **20**:1.

P.T. Strickland and J.M. Boyle (1984) Prog. Nucl. Acids Res. Mol. Biol. **31**:1.

S.A. Strobel and P.B. Dervan (1990) Science **249**:73.

R.K. Strong, R. Campbell, D.R. Rose, G.A. Petsko, J. Sharon, and M.N. Margolies (1991) Biochemistry **30**:3739.

S.W. Suh, T.N. Bhat, M.A. Navia, G.H. Cohen, D.H. Rao, S. Rudikoff, and D.R. Davies (1986) Proteins: Struct. Funct. and Genet. **1**:74.

J. Tormo, E. Stadler, T. Skern, H. Auer, O. Kanzler, C. Betzel, D. Blaas, and I. Fita (1992) Protein Science **1**:1174.

W. Trepicchio and K.J. Barrett (1987) J. Immunol. **138**:2323.

W.R. Tulip, J.N. Varghese, W.G. Laver, R.G. Webster, and P.M. Colman (1992) J. Mol. Biol. **227**:122.

L. Verlet (1967) Phys. Rev. **159**:98.

M. Weigert, R. Perry, D. Kelly, T. Hunkapillar, J. Schilling, and L. Hood (1980) Nature **283**:497.

S.J. Weiner, U.C. Singh, T.J. O'Donnell, and D.A. Kollman (1984) J. Amer. Chem. Soc. **106**:6243.

R.D. Wells, D.A. Collier, J.C. Hanvey, M. Shimizu, and F. Wohlrab (1988) FASEB J. **2**:2939.

H. Weintraub (1983) Cell **32**:1191.

S.W. White, K. Applet, K.S. Wilson, and I. Tanaka (1989) Proteins: Struct. Funct. and Genet. **5**:281.

A.F. Williams (1987) Immunol. Today **8**:298.

A.F. Williams and A.N. Barkley (1988) Annu. Rev. Immunol. **6**:381.

A.J.C. Wilson (1949) Acta Cryst. **2**:318.

T.T. Wu and E.A. Kabat (1970) J. Exp. Med. **132**:211.

Y. Yarden, J.A. Escobedo, W.J. Kuang, T.L. Yang-Feng, T.O. Daniel, P.M. Tremble, E.Y. Chen, M.E. Ando, R.N. Harkins, U. Francke, V.A. Fried, A. Ullrich, and L.T. Williams (1986) Nature **323**:226.

Ng. H. Xuong, S.T. Freer, R. Hamlin, C. Nielsen, and W. Vernon (1978) Acta Cryst. **A34**:289.

Ng. H. Xuong, D. Sullivan, C. Nielsen, and R. Hamlin (1985) Acta Cryst. **B41**:267.

Appendix A: Amino Acid Sequences of Nucleic acid Binding Antibodies.

The amino acid sequences of the variable domains of the six nucleic acid binding antibodies that are discussed in the text are presented in this appendix. The sequences are listed according to the format of Kabat *et al.* (1987), in which the Framework Regions and the Complementarity Determining Regions (CDRs) are separated by solid lines. The Kabat numbering scheme is also listed beside each sequence. This numbering scheme allows for insertions in the CDRs by numbering some of the hypervariable residues using a number followed by a letter (eg. 27a). In this way, conserved residues within a framework regions have identical residue numbers regardless of the length of the CDRs.

The residues are listed according to their three-letter amino acid code. Both the three-letter and the one-letter codes are used in the text to label residues in the figures. Listed below is the correspondance between an amino acid and its one and three-letter abbreviations.

AMINO ACID	THREE	ONE
ALANINE	ALA	A
ARGININE	ARG	R
ASPARAGINE	ASN	N
ASPARTIC ACID	ASP	D
CYSTEINE	CYS	C
GLUTAMIC ACID	GLU	E
GLUTAMINE	GLN	Q
GLYCINE	GLY	G
HISTIDINE	HIS	H
ISOLEUCINE	ILE	I
LEUCINE	LEU	L
LYSINE	LYS	K
METHIONINE	MET	M
PHENYLALANINE	PHE	F
PROLINE	PRO	P
SERINE	SER	S
THREONINE	THR	T
TYROSINE	TYR	Y
TRYPTOPHAN	TRP	W
VALINE	VAL	V

Appendix A: I. Amino Acid Sequences of the V_H Domains of Nucleic Acid Binding Antibodies

KAB Num		Jel 72	Jel 318	Hed 10	Jel 242	Jel 201	Jel 274		
1		GLU	GLU	GLU	GLN	ASP	GLU		
2		VAL	VAL	VAL	VAL	VAL	VAL		
3		GLN	GLN	LYS	GLN	GLN	GLN		
4		LEU	LEU	LEU	LEU	LEU	LEU		
5		GLN	VAL	GLU	GLN	VAL	GLN		
6		GLN	GLU	GLU	GLN	GLU	GLN		
7		SER	SER	SER	SER	SER	SER		
8		GLY	GLY	GLY	GLY	GLY	GLY		
9		GLY	GLY	GLY	THR	GLY	ALA		
10		GLU	GLY	GLY	GLU	GLY	GLU		
11		LEU	LEU	LEU	LEU	LEU	LEU		F
12		VAL	VAL	ALA	ALA	VAL	VAL		R
13		LYS	GLN	GLN	ARG	LYS	LYS		1
14		PRO	PRO	PRO	PRO	PRO	PRO		
15		GLY	LYS	GLY	GLY	GLY	GLY		
16		ALA	GLY	GLY	ALA	GLU	ALA		
17		SER	SER	SER	SER	SER	SER		
18		VAL	LEU	MET	VAL	LEU	VAL		
19		LYS	LYS	LYS	LYS	LYS	LYS		
20		LEU	LEU	LEU	LEU	LEU	LEU		
21		SER	SER	SER	SER	SER	SER		
22		CYS	CYS	CYS	CYS	CYS	CYS		
23		LYS	ALA	VAL	LYS	ALA	LYS		
24		ALA	ALA	ALA	ALA	ALA	ALA		
25		SER	SER	SER	SER	SER	SER		
26	*	GLY	GLY	GLY	GLY	ALA	GLY	*	
27		TYR	PHE	PHE	TYR	PHE	TYR		
28		THR	SER	SER	THR	THR	THR		
29		PHE	PHE	PHE	PHE	PHE	PHE		
30		THR	ASN	SER	THR	SER	THR		
31		SER	THR	ASN	SER	ASP	ASN		
32	*	TYR	TYR	TYR	TYR	TYR	SER	*	C
33		TYR	ALA	TRP	THR	GLY	TRP		D
34		MET	MET	MET	ILE	MET	ILE		R
35		TYR	ASN	ASN	THR	HIS	ASN		1
36		TRP	TRP	TRP	TRP	TRP	TRP		
37		VAL	VAL	VAL	VAL	VAL	VAL		
38		LYS	ARG	ARG	LYS	ARG	LYS		
39		GLN	GLN	GLN	GLN	GLN	GLN		

KAB Num		Jel 72	Jel 318	Hed 10	Jel 242	Jel 201	Jel 274		
40		ARG	ALA	SER	ARG	ALA	ARG		
41		PRO	PRO	PRO	PRO	PRO	PRO		F
42		GLY	GLY	GLU	GLY	GLU	GLY		R
43		GLN	LYS	LYS	GLN	LYS	GLN		2
44		GLY	SER	GLY	GLY	GLY	GLY		
45		LEU	LEU	LEU	LEU	LEU	LEU		
46		GLU	GLU	GLU	GLU	GLU	GLU		
47		TRP	TRP	TRP	TRP	TRP	TRP		
48		ILE	VAL	VAL	ILE	ILE	ILE		
49		GLY	ALA	ALA	GLY	VAL	GLY		
50		GLU	ARG	GLN	GLU	TYR	ASN		
51		ILE	THR	ILE	ILE	ILE	ILE		
52	*	ASN	MET	ARG	TYR	ASN	TYR	*	
52a		PRO	SER	LEU	PRO	SER	PRO		
52b		LYS	ARG	---	---	---	---		
52c		SER	SER	---	---	---	---		C
53		SER	LYS	ASP	ARG	GLY	GLY		D
54		ASN	ASN	ASN	SER	SER	SER		R
55	*	GLY	TYR	TYR	GLY	THR	SER	*	2
56		GLY	ALA	ALA	ASN	THR	ARG		
57		THR	THR	ILE	ALA	ILE	THR		
58		ASN	TYR	HIS	TYR	TYR	ASN		
59		PHE	TYR	TYR	TYR	TYR	TYR		
60		ASN	ALA	ALA	ILE	ALA	ASN		
61		GLU	ASP	GLU	GLU	ASP	ASP		
62		LYS	SER	SER	LYS	THR	ASN		
63		PHE	VAL	VAL	LEU	VAL	PHE		
64		LYS	LYS	LYS	ARG	LYS	LYS		
65		GLY	ASP	GLY	GLY	GLY	SER		
66		LYS	ARG	ARG	LYS	ARG	LYS		
67		ALA	PHE	PHE	ALA	PHE	ALA		
68		THR	THR	THR	THR	THR	THR		
69		LEU	ILE	ILE	LEU	ILE	LEU		
70		THR	SER	SER	THR	SER	THR		
71		VAL	ARG	ARG	VAL	ARG	VAL		
72		ASP	ASP	ASP	ASP	ASP	ASP		
73		LYS	ASP	ASP	LYS	ASN	THR		
74		SER	SER	SER	SER	ALA	SER		
75		SER	GLN	LYS	SER	LYS	SER		
76		SER	SER	SER	ASN	ASN	SER		
77		ILE	MET	SER	THR	THR	THR		F
78		ALA	LEU	VAL	ALA	LEU	ALA		W
79		TYR	TYR	TYR	TYR	PHE	TYR		2
80		MET	LEU	LEU	MET	LEU	MET		
81		GLN	GLN	GLN	GLN	GLN	GLN		

KAB Num		Jel 72	Jel 318	Hed 10	Jel 242	Jel 201	Jel 274		
82		LEU	MET	MET	LEU	MET	LEU		
82A		SER	ASN	ASN	SER	THR	SER		
82B		SER	ASN	ASN	SER	SER	SER		
82C		LEU	LEU	LEU	LEU	LEU	LEU		
83		THR	LYS	ARG	LYS	ARG	THR		
84		SER	THR	ALA	SER	SER	SER		
85		GLU	GLU	GLU	GLU	GLU	ASP		
86		ASP	ASP	ASP	ASP	ASP	ASP		
87		SER	ARG	THR	SER	THR	SER		
88		ALA	ALA	ALA	ALA	ALA	ALA		
89		VAL	ILE	ILE	VAL	MET	VAL		
90		TYR	TYR	TYR	TYR	TYR	TYR		
91		TYR	TYR	TYR	PHE	TYR	TYR		
92		CYS	CYS	CYS	CYS	CYS	CYS		
93		THR	VAL	THR	ALA	ALA	ALA		
94		LYS	ARG	VAL	ARG	ARG	ARG		
95		GLY	GLU	TYR	TYR	ASN	TRP		
96	*	GLY	LEU	TYR	GLY	ASN	ARG	*	
97		SER	LEU	GLY	ASN	TYR	ASP		
98		ARG	ARG	TYR	TYR	TYR	TYR		
99		VAL	SER	LEU	VAL	GLY	ARG		
100		ARG	PHE	---	ARG	SER	SER		C
100a		ARG	---	---	PHE	SER	PHE		D
100b		TYR	---	---	---	PRO	PHE		R
100c		TYR	---	---	---	PHE	---		3
100d		ALA	---	---	---	---	---		
100e		MET	---	---	---	---	---		
101	*	ASP	ALA	GLY	ASP	ALA	ALA	*	
102		TYR	TYR	PHE	TYR	TYR	TYR		
103		TRP	TRP	TRP	TRP	TRP	TRP		
104		GLY	GLY	GLY	GLY	GLY	GLY		
105		GLN	GLN	GLN	GLN	GLN	GLN		F
106		GLY	GLY	GLY	GLY	GLY	GLY		W
107		THR	THR	THR	THR	THR	THR		3
108		LEU	LEU	THR	THR	LEU	LEU		

The Framework and Complementarity Determining Regions, as determined from a comparison of antibody sequences, (Kabat et al.,1987) are as indicated.

The Hypervariable Regions as determined from a comparison of known Fab fragment structures (Chothia and Lesk, 1987) marked by an asterisk (*).

Appendix A: II. Amino Acid Sequences of the V_L Domains of Nucleic Acid Binding Antibodies

KAB Num	Jel 72	Jel 318	Jel 242	HED 10	Jel 201	Jel 274	
1	ASP	ASP	ASP	ASP	ASP	ASP	
2	VAL	ILE	ILE	ILE	ILE	ILE	
3	VAL	VAL	GLN	VAL	VAL	VAL	
4	MET	LEU	MET	MET	MET	MET	
5	THR	THR	SER	THR	THR	THR	
6	GLN	GLN	GLN	GLN	GLN	GLN	
7	THR	SER	SER	ALA	ALA	THR	
8	PRO	PRO	THR	ALA	ALA	PRO	
9	LEU	ALA	SER	PRO	PRO	LEU	
10	THR	ILE	SER	SER	SER	ILE	F
11	LEU	MET	LEU	VAL	VAL	LEU	R
12	SER	SER	SER	PRO	PRO	SER	1
13	VAL	ALA	ALA	VAL	ASP	VAL	
14	THR	SER	SER	THR	THR	THR	
15	ILE	PRO	LEU	PRO	PRO	ILE	
16	GLY	GLY	GLY	GLY	GLY	GLY	
17	GLN	GLU	ASP	GLU	GLU	GLN	
18	PRO	LYS	ARG	SER	SER	PRO	
19	ALA	VAL	VAL	VAL	VAL	ALA	
20	SER	THR	THR	SER	SER	SER	
21	ILE	ILE	ILE	ILE	ILE	ILE	
22	SER	THR	SER	SER	SER	SER	
23	CYS	CYS	CYS	CYS	CYS	CYS	
24	LYS	SER	ARG	ARG	ARG	LYS	
25	SER	ALA	ALA	SER	SER	SER	
26 *	SER	SER	SER	SER	SER	SER	*
27	GLN	---	GLN	THR	LYS	GLN	C
27a	SER	---	---	SER	SER	SER	D
27b	LEU	---	---	LEU	LEU	LEU	R
27c	LEU	---	---	LEU	LEU	LEU	1
27d	ASP	---	---	HIS	HIS	TYR	
27e	SER	---	---	SER	SER	SER	
28	ASP	SER	ASP	SER	ASN	ASN	
29	GLY	SER	ILE	GLY	GLY	GLY	
30	LYS	VAL	SER	LYS	ASN	LYS	
31	THR	SER	ASN	ASN	THR	THR	
32 *	TYR	TYR	ASN	ARG	TYR	TYR	*
33	LEU	MET	LEU	LEU	LEU	LEU	
34	ASN	HIS	ASN	TYR	TYR	ASN	
35	TRP	TRP	TRP	TRP	TRP	TRP	
36	LEU	TYR	TYR	PHE	PHE	LEU	
37	LEU	HIS	ARG	LEU	LEU	LEU	

KAB Num	Jel 72	Jel 318	Jel 242	HED 10	Jel 201	Jel 274	
38	GLN	GLN	GLN	GLN	GLN	GLN	
39	ARG	LYS	LYS	ARG	ARG	ARG	
40	PRO	SER	PRO	PRO	PRO	PRO	F
41	GLY	GLY	ASP	GLY	GLY	GLY	R
42	GLN	THR	GLY	GLN	GLN	GLN	2
43	SER	SER	THR	SER	SER	SER	
44	PRO	PRO	VAL	PRO	PRO	PRO	
45	LYS	LYS	LYS	GLN	HIS	LYS	
46	ARG	ARG	LEU	LEU	LEU	ARG	
47	LEU	TRP	LEU	LEU	LEU	LEU	
48	ILE	ILE	ILE	ILE	ILE	ILE	
49	TYR	TYR	TYR	TYR	TYR	TYR	
50 *	LEU	ASP	TYR	TYR	ARG	LEU	*
51	VAL	THR	THR	MET	MET	VAL	C
52 *	SER	SER	SER	SER	SER	SER	* D
53	LYS	LYS	ARG	ASN	ASN	LYS	R
54	LEU	LEU	LEU	LEU	LEU	LEU	2
55	ASP	ALA	HIS	ALA	ALA	ASP	
56	SER	SER	SER	SER	SER	SER	
57	GLY	GLY	GLY	GLY	GLY	GLY	
58	VAL	VAL	VAL	VAL	VAL	VAL	
59	PRO	PRO	PRO	PRO	PRO	PRO	
60	ASP	ALA	SER	ASP	ASP	ASP	
61	ARG	ARG	ARG	ARG	SER	ARG	
62	PHE	PHE	PHE	PHE	PHE	PHE	
63	THR	SER	SER	SER	SER	THR	
64	GLY	GLY	GLY	GLY	GLY	GLY	
65	SER	SER	SER	SER	SER	SER	
66	GLY	GLY	GLY	GLY	GLY	GLY	
67	SER	SER	SER	SER	SER	SER	
68	GLY	GLY	GLY	GLY	GLY	GLY	
69	THR	THR	THR	THR	THR	THR	F
70	ASP	SER	ASP	ALA	ALA	ASP	R
71	PHE	TYR	TYR	PHE	PHE	PHE	3
72	THR	SER	SER	THR	THR	THR	
73	LEU	LEU	LEU	LEU	LEU	LEU	
74	LYS	THR	THR	ARG	ARG	LYS	
75	ILE	ILE	ILE	ILE	ILE	ILE	
76	SER	SER	SER	SER	SER	SER	
77	ARG	SER	ASN	ARG	ARG	ARG	
78	VAL	MET	LEU	VAL	VAL	VAL	
79	GLU	GLU	GLU	GLU	GLU	GLU	
80	ALA	ALA	GLN	ALA	ALA	ALA	
81	GLU	GLU	GLU	GLU	GLU	GLU	
82	ASP	ASP	HIS	GLY	ASP	ASP	

KAB Num		Jel 72	Jel 318	Jel 242	HED 10	Jel 201	Jel 274	
83		PHE	ALA	ILE	PHE	VAL	LEU	
84		GLY	SER	ALA	GLY	ARG	GLY	
85		VAL	THR	THR	ALA	ALA	VAL	
86		TYR	TYR	TYR	TYR	TYR	TYR	
87		TYR	TYR	TYR	TYR	TYR	TYR	
88		CYS	CYS	CYS	CYS	CYS	CYS	
89		TRP	GLN	GLN	MET	MET	VAL	
90		GLN	GLN	GLU	GLN	GLN	GLN	
91	*	GLY	LEU	GLY	SER	HIS	GLY	*
92		THR	SER	ASN	LEU	LEU	THR	C
93		HIS	SER	THR	GLN	GLU	HIS	D
94		PHE	ASN	LEU	TYR	TYR	PHE	R
95		PRO	PRO	PRO	PRO	PRO	PRO	3
96	*	GLN	TYR	ARG	TYR	TYR	TYR	*
97		THR	THR	THR	THR	THR	TYR	
98		PHE	PHE	PHE	PHE	PHE	PHE	
99		GLY	GLY	GLY	GLY	GLY	GLY	
100		GLY	GLY	GLY	GLY	GLY	GLY	F
101		GLY	GLY	GLY	GLY	GLY	GLY	R
102		THR	THR	THR	THR	THR	THR	4
103		LYS	LYS	LYS	LYS	LYS	LYS	
104		LEU	LEU	ILE	LEU	LEU	LEU	
105		GLU	GLU	GLU	GLU	GLU	GLU	
106		ILE	ILE	ILE	ILE	ILE	ILE	
107		LYS	LYS	LYS	LYS	LYS	LYS	

The Framework and Complementarity Determining Regions, as determined from a comparison of antibody sequences, (Kabat et al., 1987) are as indicated.

Thr Hypervariable Regions as determined from a comparison of known Fab fragment structures (Chothia and Lesk, 1987) are marked by ab asterisk (*).